

Do partially disabled people respond to financial incentives to work?

Wietse Mesman*, Tunga Kantarci† and Jan-Maarten van Sonsbeek‡

12 January 2021

Preliminary version

Abstract

We evaluate the work resumption program of the Dutch disability insurance (DI) scheme for partially disabled individuals. Participants face financial incentives to increase labor supply which are stronger if pre-sickness wage is higher. Taking a local randomization approach to regression discontinuity and using administrative data on the universe of participants, we find clear heterogeneous responses to the program. Individuals with low pre-sickness wages do not respond whereas labor participation increases among those with higher pre-sickness wages. Moreover, individuals with higher pre-sickness wages continue to increase labor participation during the two years after the work resumption program takes effect whereas those with low pre-sickness wages remain unresponsive.

1 Introduction

In 2017, the Netherlands spent 2.9% of its GDP on sickness and disability benefits. Although still well above the OECD average of 2.0% (in similar years), public spending on incapacity in the Netherlands is now markedly lower than in the Scandinavian countries that top the OECD list with, on average, over 4% of their GDP ([OECD, 2021](#)). Despite an aging population, spending on sickness and disability has decreased in the Netherlands since the beginning of the century as a result of a series of reforms.

Figure 1 presents the stock of disability benefit recipients and inflow into the Dutch DI scheme as fractions of the insured population during the period from 1969 to 2021. The stock of disability benefit recipients reached an all-time high in 1990 leading to a first set of major sickness insurance (SI) and DI reforms in that decade. Although initially successful, by the end of the century it became clear that these reforms did not succeed in structurally reducing the number of DI recipients.

*Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, and Netspar (e-mail: w.r.b.mesman@tilburguniversity.edu)

†Department of Econometrics and Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, and Netspar (e-mail: kantarci@tilburguniversity.edu)

‡Department of Public Finance, Netherlands Bureau for Economic Policy Analysis, P.O. Box 80510, 2508 GM The Hague, The Netherlands, and Netspar (e-mail: j.m.van.sonsbeek@cpb.nl)

Consequently, DI reform appeared on the political agenda again, and an expert committee was set up to advise on the reforms. This committee drew the contours of a new law in 2001, proposing a paradigm shift, changing the focus from “what people are not able to do anymore” to “what people are still able to do”. In their original proposal, DI was limited to the permanently fully disabled, whereas all temporarily fully disabled and partially disabled could rely on the less generous and temporary unemployment insurance (UI) benefits. The main lines of the committee’s proposal, among them stricter eligibility criteria and an extended sickness period, were followed in the new Work and Income According to Labor Capacity Act (WIA), but their treatment of the partially disabled was considered too strict. Therefore, a new scheme aimed at the temporarily fully disabled and partially disabled, later to be known as the Work resumption program for partially and temporarily fully disabled (WGA), was added to the new law. This scheme was designed to initially mimic the UI scheme, and thereafter provide strong financial incentives so that the partially disabled are stimulated to use their remaining work capacity.

While the new WIA law was being developed, other reforms were already implemented. In 1998 financial incentives for employers were introduced by means of an experience rated DI premium, that was stepwise increased until 2002. Also in 2002, the “Gatekeeper protocol” was introduced, in which clear and concrete mutual reintegration obligations of employers and sick employees during the sickness period are specified. In January 2004 the sickness period was extended from 1 to 2 years. Finally, in 2006, the WIA was introduced, effectively covering the workers who reported sick from 2004 onwards. These reforms are considered to be effective and responsible for the large decreases of new disability benefit awards since their introduction ([Van Sonsbeek and Gradus, 2013](#); [Koning and Lindeboom, 2015](#); [Kantarci et al., 2019](#)).

In recent years, however, inflow into the DI scheme is increasing again. This can be explained by the increasing labor participation of older workers, as early retirement schemes have been phased out in the last two decades, and the statutory retirement age was increased in 2013. The COVID pandemic might have contributed to the pronounced increase in 2020 and 2021. Both factors may complicate reintegration efforts and cause (additional) health problems.

In the WIA, sick employees first stay for two years with their employer before they can claim a DI benefit. Thereafter, partially disabled individuals qualify for a DI benefit that replaces 70% of the pre-sickness wage for a period of 3 to a maximum of 38 months. The duration is equal to the duration of UI benefit because in fact UI benefit is incorporated in the DI benefit. After this period, they face a work resumption program that provides strong financial incentives. The UI benefit component of the DI benefit expires and no longer constitutes part of the DI benefit. Furthermore, if the partially disabled do not work a threshold number of hours, their benefit drops to a substantially lower level.

Although the financial incentives of the work resumption program are strong, they are offered only after a period of sickness of up to 5 years and 2 months, thereby complicating a successful return to the labor market. Indeed, work resumption from disability may be easiest when this happens with the former employer and as soon as possible after falling sick ([Koning and Lindeboom, 2015](#)). If without a job when subjected to the program incentives, individuals may struggle to resume working if job search costs are high, negotiating a suitable work schedule with an employer is difficult, or adjusting non-work schedules is difficult ([Zaresani, 2020](#)). They may also find that the market wage in the new job is below their reservation wage.

Besides the possible labor market restrictions, by design, financial incentives of the work resumption program are weaker for those who earn lower wages before they fall sick. For them, the DI benefit is smaller in absolute terms but also relative to pre-sickness earnings, making work resumption less attractive.

Recent studies analyze the effects of financial incentives to increase labor participation

among disabled individuals. Most of these studies exploit exogenous changes in benefit rules from policy reforms or field experiments to investigate the impact of the financial incentives. [Campolieti and Riddell \(2012\)](#) analyze the effect of the change in the amount of wages that DI beneficiaries are allowed to earn without losing their disability benefits. They find a substantial increase in labor participation of beneficiaries. [Kostøl and Mogstad \(2014\)](#) study a reform that abolished discontinuous penalties on benefits due to earning above threshold wages in Norway. They find a substantial positive effect on labor participation. [Vall Castelló \(2017\)](#) analyzes the removal of the tax exemption for unemployed partial DI beneficiaries younger than 55 in Spain. She finds that the exemption increased labor participation of the affected group by 6.5 percentage points (pp). In a field experiment in the US, [Weathers and Hemmeter \(2011\)](#) find that the number of beneficiaries with earnings above the threshold increases by 25% when they lose 1 dollar for every 2 dollars in additional earnings instead of when they lose the entire benefit above the threshold. [Bütler et al. \(2015\)](#) study a field experiment in Switzerland. Randomly selected DI benefit recipients were offered a lump-sum transfer on the condition that they increase their labor supply and reduce benefit claims. By the end of the three-year program, only 0.5% took up the offer.

To the best of our knowledge, only two studies investigate the impact of existing financial incentives built into the DI schemes. Many countries reduce DI benefits if earnings exceed a threshold amount. This provides a “cash cliff” that benefit recipients may try to avoid by bunching their earnings just below the threshold. In Austria, [Ruh and Staubli \(2019\)](#) estimate that the average DI recipient with earnings just below the threshold would increase earnings by 45% in the absence of the threshold.

In the Netherlands, [Koning and van Sonsbeek \(2017\)](#) analyze the effects of the work resumption program built in the DI scheme on labor participation, daily earnings, and full work resumption. To identify causal effects, they exploit the duration of the disability benefit that workers are entitled until they face the financial incentives of the work resumption program. Duration of this benefit depends on the work history of an individual which induces substantial variation across the study sample. Comparing the workers close to or subjected to the incentives to those who are not (yet) subjected, they identify a short-term effect in the six months around the individual-specific dates financial incentives take effect, and a long-term effect for the period afterwards. They find that the short- and long-term increases in labor participation are 1.4 and 2.6 pp, respectively. Studying the heterogeneous effects of the program, among others, they find that individuals with pre-sickness daily wages both below and above €125 increase labor participation by 1.8 and 2.3 pp, respectively. They find a limited impact on wages and no significant impact on full work resumption.

Like [Koning and van Sonsbeek \(2017\)](#), we investigate the labor supply effects of the work resumption program of the Dutch DI scheme. Our study, however, differs from theirs in three respects. First, we focus on the heterogeneous incentives of the work resumption program across income groups. We recognize that, by design, financial incentives of the work resumption program are weaker for those who earn lower wages before they fall sick. This makes work resumption less attractive for the comparatively large group of low-income earners. Disability benefits are already near the minimum wage for a large part of this group, and possibilities for increasing their financial incentives for work seem limited. The opportunities to increase their labor participation may be limited not only by their disability but also lack of skills. Second, to identify causal effects, we take a local randomization approach to regression discontinuity as a new and transparent identification strategy. We conduct falsification tests that identify the regions where identifying assumptions are most likely to hold. Third, we analyze whether financial incentives work better or worse at different stages of the business cycle. How financial incentives interact with the business cycle is not evident because effects of the business cycle on

the behavior of employers and workers with disabilities may be different. For example, during downturns of the business cycle, employers may find workers with disabilities less attractive, whereas workers with disabilities may try harder to find suitable jobs. We use the business cycle indicator of the Dutch Central Bank to identify the impact of the changes in the Dutch business cycle.

We find immediate positive effects of the work resumption program on labor participation for two groups with higher pre-sickness wages. For three groups with higher pre-sickness wages, the coefficient estimates increase notably in the first few months. Thereafter they increase more gradually or stabilize. In contrast, we find little evidence of an effect for lower pre-sickness wage groups. These results are in line with the design of the work resumption program which provides no or small incentives to increase labor supply for individuals with pre-sickness wages below or around the minimum wage.

The remainder of this paper is structured as follows. Section 2 describes the Dutch DI scheme for partially disabled individuals. Section 3 introduces the data and Section 4 presents descriptive statistics. Section 5 describes the identification strategy. Section 6 presents the estimation results. Section 7 conducts falsification and sensitivity tests. Section 8 concludes with policy suggestions.

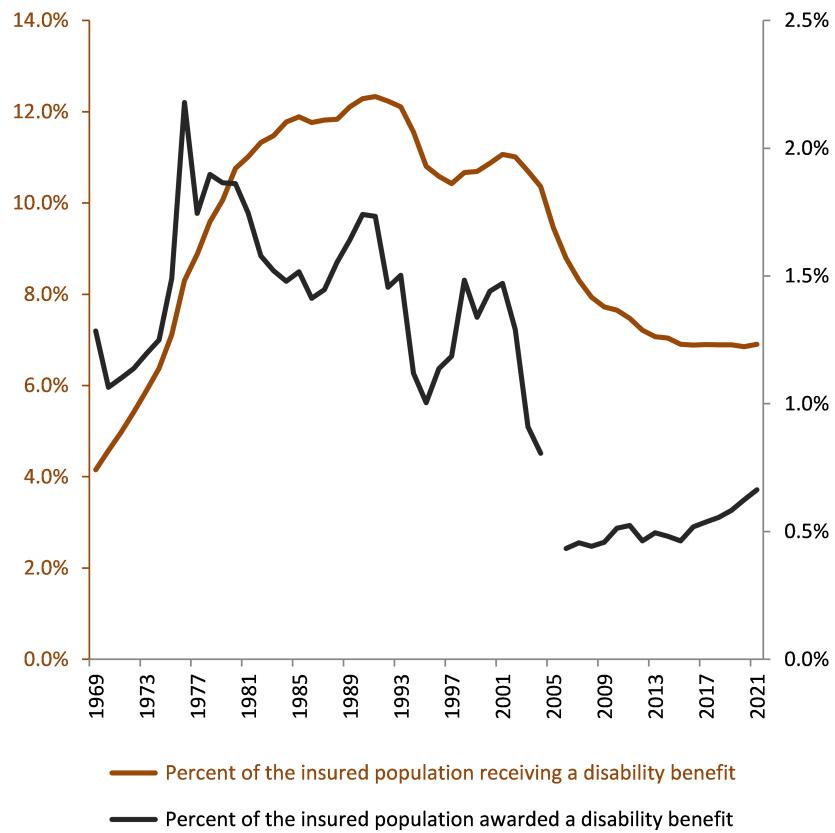


Figure 1: Fractions of the insured population receiving and awarded disability benefits during the period from 1969 to 2021. No disability benefits are awarded in 2005 due to the DI reform in 2004 that extended the duration of the SI scheme, which precedes the DI scheme, from one to two years. Source: The Employee Insurance Agency (UWV) and Statistics Netherlands.

2 The Dutch disability insurance scheme and the work resumption program

The WIA came into effect on 1 January 2006 for people who fell ill from 1 January 2004 onwards. In the WIA, workers who lose any part of their earning capacity due to a health impairment are entitled to the sickness benefit from their employer for a period of two years. When the sickness benefit expires, the worker can apply for a disability benefit. If the wage loss of the individual is more than 80%, with no possibility of recovery, the worker is admitted to the Income Provision Scheme for Fully Occupationally Disabled People (IVA). We do not study the labor supply responses of IVA beneficiaries, since their remaining earning capacity is very limited and there is hardly any possibility of work resumption. If the wage loss is more than 35% but less than 80%, or if the wage loss is more than 80% but there is a possibility of recovery, the worker is admitted to the Return to Work Scheme for the Partially Disabled (WGA). In the current study we focus on this scheme.

The WGA consists of two stages. The worker is first entitled to the “wage-related benefit”. This benefit consists of two parts. The first part is fixed and equal to 70% of the pre-sickness monthly wage multiplied by the disability grade.¹ The second part is variable and equal to 70% of the pre-sickness wage multiplied by (1 - disability grade) and (1 - utilization rate), that is, the unutilized part of the remaining earning capacity.^{2,3} The second part is an unemployment benefit component, which compensates individuals who are not able to utilize their remaining work capacity. Beneficiaries who utilize (part of) their remaining earning capacity see their benefits reduced by € 0,70 for every € 1 of wage income. Thus, in this part of the scheme there is a high implicit tax rate on wages. Duration of the wage-related benefit depends on the person’s employment history, and it is limited to a maximum of 38 months.

When the wage-related benefit expires, individuals are entitled to one of two types of benefits depending on whether they utilize at least 50% of their remaining earnings capacity. If they utilize at least 50% of remaining earning capacity, they are entitled to the “wage-supplement benefit”, which replaces 70% of the pre-sickness wage multiplied by the disability grade. If they utilize less than 50% of remaining earning capacity, they are entitled to the “follow-up benefit”, which replaces 70% of the pre-sickness wage multiplied by the disability degree, but the pre-sickness wage is capped at the minimum wage.⁴

These rules imply that both the wage-supplement and follow-up benefits make flat-rate payments and therefore disregard how much the individual is working below or above the threshold utilization rate of remaining work capacity. Both benefits are paid as long as the individual is disabled but expire when the individual becomes entitled to the state pension at the statutory retirement age.

These benefit rules imply two financial incentives to increase work effort when the first-stage wage-related benefit expires. The first incentive is due to the reduction in the amount of the disability benefit as the unemployment benefit built into the first-stage wage-related benefit expires, and is no longer a component of the follow-up benefit or the wage-supplement benefit

¹That is, the amount is fixed as long as the disability grade is not reassessed.

²During the first two months of disability the fractions are 75% instead of 70%.

³It is possible for the utilization rate to be larger than one if remaining earning capacity is incorrectly estimated. In this case, the second part of the benefit is negative. If this situation persists it may lead to a reassessment of the disability grade.

⁴Beneficiaries of a DI benefit receive decision letters from the Employee Insurance Agency on two occasions. First, they receive a letter on their application for the wage-related benefit at the end of their participation in the SI scheme. Second, they receive a letter approximately three months before their wage-related benefit expires informing them of the entitlement decision of a second-stage benefit. Figures 15a to 15e and Figures 16a to 16c in the Appendix show anonymized letters of the two kinds sent to a client.

in the second stage of the DI scheme. The second incentive is to utilize at least 50% of the remaining work capacity, since in the second stage of the scheme the wage-supplement benefit is higher than the follow-up benefit.

Figure 2a illustrates the financial incentives of the work resumption program. For a given pre-sickness wage amount, the figure shows how daily income changes with the amount of work, where the latter is measured in terms of the daily wage during disability as a fraction of the pre-sickness daily wage. Daily income is from earnings and a disability benefit. During disability, earned daily wage that is 25% of the pre-sickness daily wage corresponds to a remaining work capacity utilization rate of 50% with an assumed disability grade of 50%. If workers utilize less than 50% of their remaining work capacity, the figure shows the follow-up benefit amount received in the second stage, alongside the first-stage wage-related benefit, earnings, and the total of a disability benefit and earnings. If they utilize at least 50% of their remaining work capacity, the figure shows the wage-supplement benefit received in the second stage, alongside the first-stage wage-related benefit, earnings, and the total of a disability benefit and earnings. A comparison of the total income in the first stage of the DI scheme to that in the second stage shows the extent of the financial incentive to increase labor supply in the second stage of the DI scheme.

There are a number of notable patterns. First, disability benefits are always lower in the second stage of the scheme, unless individuals utilize their remaining work capacity to the full extent, since the unemployment benefit component of the first-stage wage-related benefit expired. Second, people who utilize less than 50% of their remaining earning capacity (and qualify for the lower follow-up benefit in the second stage) face a larger penalty than those who utilize at least 50% (and qualify for the higher wage-supplement benefit in the second stage). Third, people who work more face smaller penalties. Finally, individuals with higher pre-sickness wages face larger penalties; their benefit decreases by a larger amount from the first to the second stage of the scheme. This is illustrated in Figure 2b which shows the program incentives when pre-sickness wage amount is smaller.

The penalties for individuals with high pre-sickness wages are not only larger than the penalties for those with low pre-sickness wages in absolute terms, but also relative to pre-sickness wages. Although the loss of the unemployment component of the wage-related benefit is proportional to the pre-sickness wage, the penalty for not meeting the threshold remaining earning capacity utilization rate of 50% is based on the difference between the pre-sickness wage and pre-sickness wage capped by the minimum wage. For pre-sickness wages below the minimum wage this difference is non-existent. Above that, the difference increases with the pre-sickness wage.

The financial incentive may depend on whether the individual is entitled to the social minimum supplement. For some individuals the follow-up benefit is so low that they may qualify for the supplement, which compensates part of the reduction in benefit amounts.⁵ Ceteris paribus, this case is more likely to apply to individuals with low pre-sickness wages, reducing their incentives even more. What is more, for individuals for whom the social minimum is higher than 70% of their pre-sickness wage, all benefits are below the social minimum and therefore there is no difference in income between being unemployed in the first and second stages of the WGA. All in all, for partial DI beneficiaries with low pre-sickness wages it is generally not possible to reduce income between the first and second stages by a large amount, or they would not be able to meet their ends.

⁵If an individual has a partner with an income, they may get less or no social minimum supplement. Thus not everyone with very low (own) income qualifies for the supplement.

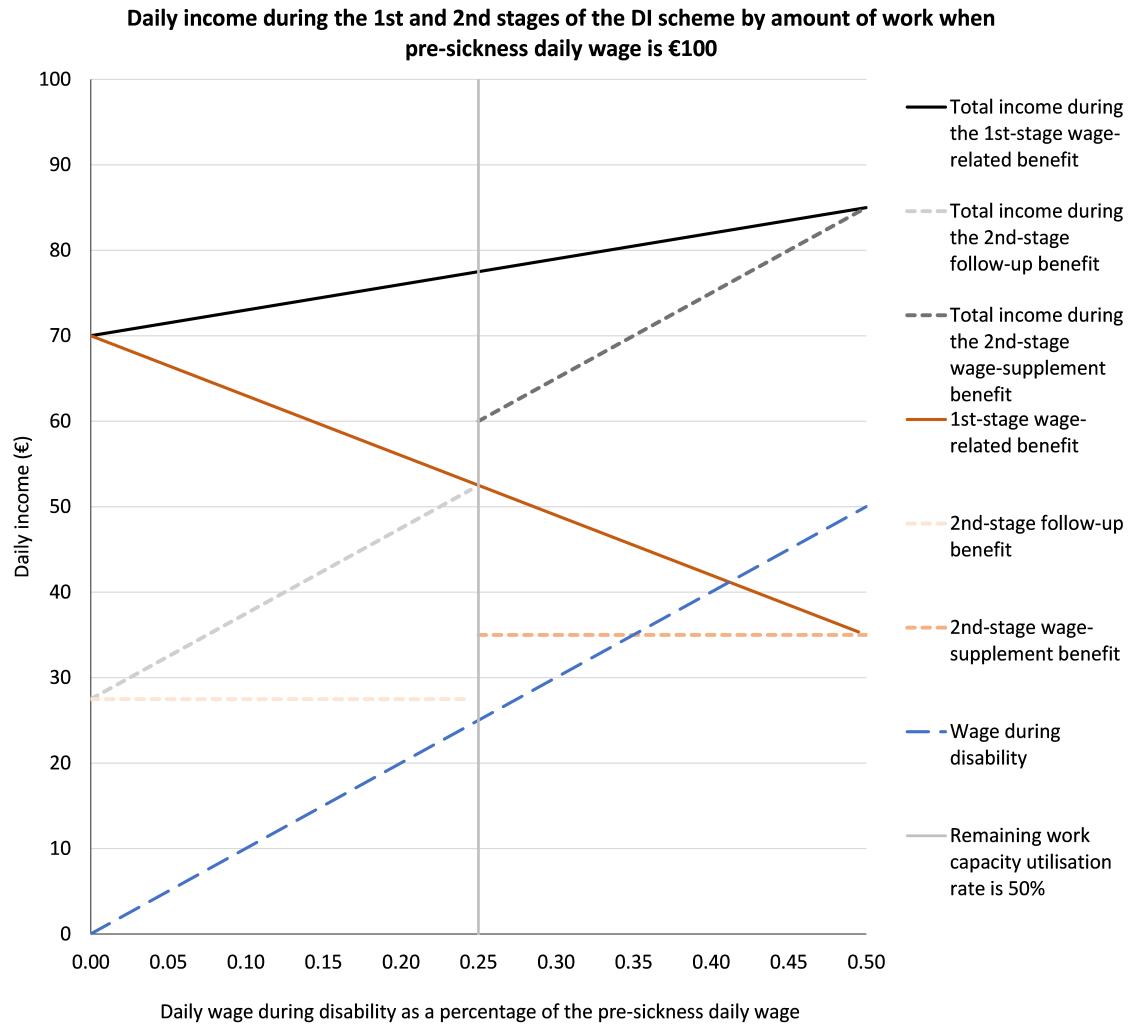


Figure 2a: Work incentives in the WGA. If workers utilize less than 50% of their remaining work capacity (left side of solid gray line), the financial incentive is the difference between the total income from wages and the wage-related benefit (solid black line) in the first stage and the total income from wages and the follow-up benefit (dashed light gray line) in the second stage. If workers utilize at least 50% of their remaining work capacity (right side of solid gray line), the financial incentive is the difference between the total income from wages and the wage-related benefit (solid black line) in the first stage and the total income from wages and the wage-supplement benefit (dashed dark gray line) in the second stage. The figure considers a pre-sickness daily wage amount of €100, and disregards the social minimum supplement.

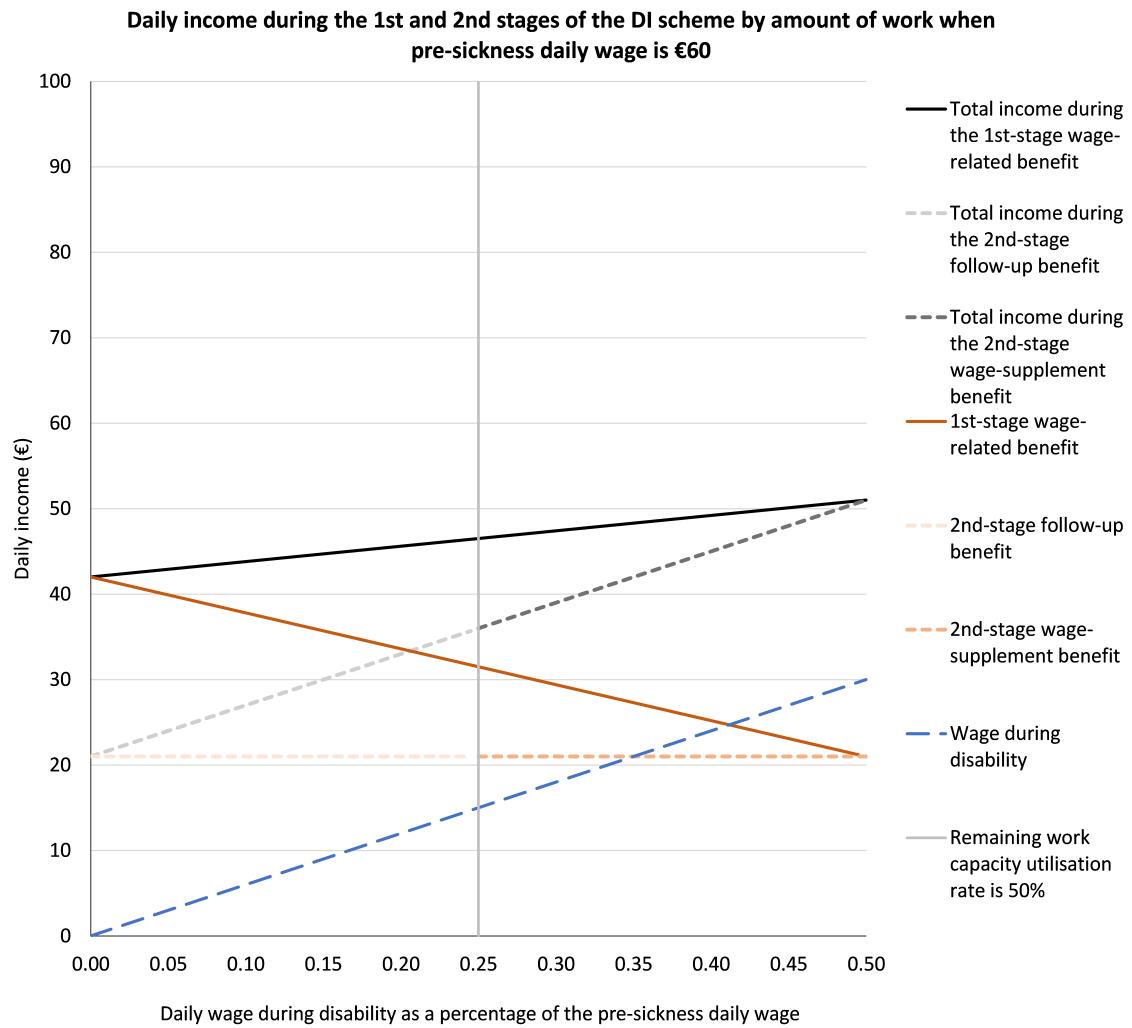


Figure 2b: Work incentives in the WGA. If workers utilize less than 50% of their remaining work capacity (left side of solid gray line), the financial incentive is the difference between the total income from wages and the wage-related benefit (solid black line) in the first stage and the total income from wages and the follow-up benefit (dashed light gray line) in the second stage. If workers utilize at least 50% of their remaining work capacity (right side of solid gray line), the financial incentive is the difference between the total income from wages and the wage-related benefit (solid black line) in the first stage and the total income from wages and the wage-supplement benefit (dashed dark gray line) in the second stage. The figure considers a pre-sickness daily wage amount of €60 and disregards the social minimum supplement.

3 Data

We use administrative data provided by the Employee Insurance Agency on the universe of individuals who participate in the DI scheme for partially disabled workers (WGA) during the observation period. The data includes monthly information on the type of the DI benefit received (wage-related, wage-supplement, and follow-up benefits), disability grade, remaining earning capacity, wages earned before reporting sick and during disability, and a limited number individual characteristics such as age and gender. Labor participation during DI receipt, as the outcome variable, is defined as having positive earnings in a month. The observation period starts in January 2006, when sick individuals have become eligible for DI benefits the first time in the WIA, and ends in December 2013. Only individuals who enter the DI scheme until 2010 are included in the analysis. They are the beneficiaries who are observed from the moment they enter the scheme in 2006 up until the moment they exit again or the observation period ends.

The original data includes individuals who re-enter the DI scheme after exiting, move between the first and second stages of the DI scheme multiple times, or are not observed in some months. Individuals with such irregular trajectories are excluded from the analysis. Furthermore, individuals who enter the DI scheme for fully disabled people (IVA) at any point in time are excluded. Even if some of them face the financial incentives, they are fundamentally different in their ability to work and therefore to respond to the incentives. The incentives are not aimed at people with little or no ability to work, and their effects on them is not the focus of this study. Finally, we exclude individuals if they exit the scheme during the first stage or do not reach the second stage at the end of the observation period since they do not experience the financial incentives. These restrictions lead to the study sample that consists of 633,251 monthly observations for 11,366 individuals.

4 Time trends and other descriptive statistics

Figure 3 shows average labor participation during a period of four years around the time the work resumption program takes effect, that is, when the first stage of the DI scheme for partially disabled individuals expires. We distinguish among five groups based on pre-sickness wage levels, defined in relative terms to the minimum wage. In the beginning of the period average labor participation of the five income groups are close to each other and fall within a range of 50% to 58%. By the end of the period, however, labor participation of the highest three and lower income groups end up being very different, ranging from about 35% to 65%. For the two lowest income groups, labor participation decreases substantially during the two years until the work resumption program takes effect. It remains fairly stable after this period. By contrast, labor participation of higher income groups is stable until about when the work resumption program starts, and increases markedly thereafter. As explained in Sections 1 and 2, in our analyses we distinguish among pre-sickness wage groups because financial incentives of the work resumption program are different for them. Considering the first stage of the DI scheme, the very different labor participation rates of income groups observed in the figure contributes to the relevance of distinguishing among income groups to analyze their subsequent labor participation in the second-stage of the DI scheme. Apparently, labor market attachment is different across income groups during the first stage of the DI scheme.

It is important to note, however, that changes in the group averages are not driven solely by individuals in the groups changing their labor participation status, but also by changes in the composition of the groups. Pre-sickness wage is fixed, meaning that individuals do not move between income groups, but they may enter the sample less than 24 months before the program incentive takes effect and/or exit the sample less than 24 months afterwards. Figure 4 illustrates

this.⁶ Everyone in the sample is observed in the last month of the first stage and the first month of the second stage. However, the further away from the moment that the financial incentive takes effect, the fewer individuals there are. In the data we observe that individuals with shorter durations of the first-stage wage-related benefit have lower labor participation. This means that the labor participation trends in the figure reflect to some extent a composition effect.

Table 1 presents sample means of background characteristics and labor market outcomes for select pre-sickness wage groups. The means are for two months, before and after the month individuals face the work resumption program. There are clear differences across the wage groups. Individuals with high pre-sickness wages are older, more often male, and more often have a higher disability grade.⁷ They also have a longer duration of the first stage, that is, a longer work history. During the second stage, they receive the wage-supplement benefit more often than the follow-up benefit, and the difference increases with pre-sickness wage. On the other hand, individuals in the lowest pre-sickness wage group spend almost twice as long receiving the follow-up benefit than the higher wage-supplement benefit. This is in line with their considerably lower labor participation since higher labor participation implies a higher probability of receiving the wage-supplement benefit (Section 2). Across the groups with lower wages, much larger shares of people receive the social minimum supplement to increase their income up to the social minimum.

Although the sample size of the lowest pre-sickness wage group is considerably smaller than the sample sizes of all other groups, within each group the samples are the same in the last month of the first stage and the first month of the second stage. Therefore, differences within groups in these two months are not due to changes in the composition of the groups. Hence, the gender (im)balance is constant across the two periods.⁸ While marital status and disability grade can change over time, large changes are not expected in consecutive months, and the figures in the table are in line with this expectation.⁹ Age necessarily increases by one-twelfth of a year with each month spent in the DI scheme. As argued above, receiving the social minimum supplement is endogenous to the treatment and therefore the significant differences in this variable do not provide evidence of a covariate imbalance before and after the program incentive. For all pre-sickness wage groups, there is an increase in labor participation following the financial incentive. Wages earned during disability also show an increase. This is due in part to the increase in labor participation, and in part to an increase in wages among individuals who were already working.

Figure 5 presents labor participation against duration of DI receipt, age, and calendar time. For all wage groups, the left panel shows a strong positive relationship between labor participation and duration of DI receipt in the first 40 months. This can be explained by

⁶The distribution has a lower end at 60 months which corresponds to the maximum duration of the first stage for individuals who entered the scheme in 2006 and 2007. The maximum duration is 38 months for individuals who entered the scheme in 2008 or later due to a policy reform in this year. This explains the smaller number of individuals with a duration of less than 38 months in the figure. The higher end of the distribution is limited by the date individuals are last observed in the data which corresponds to 92 months.

⁷Disability grade is defined as the difference between wage of a reference person and assessed remaining earning capacity, divided by the reference person wage. The reference person wage is the theoretical wage of someone who is similar to the partially disabled individual in all respects except that they have no disability. Therefore, it likely has a strong correlation with pre-sickness wage. The remaining earning capacity is based on a set of jobs that the partially disabled person can still do. For those with low pre-sickness wages (and therefore likely low wages of a reference person), there will not be any jobs with hourly wages far below their pre-sickness wage, as the minimum wage provides a lower bound. For individuals with high pre-sickness wages, there is more room for a remaining earning capacity considerably below their reference income.

⁸In principle, the same should hold for real pre-sickness daily wage. However, in some instances there are changes in the nominal values which cannot be explained by indexation. In addition, there may be rounding error in the derivation of the real values.

⁹This is despite a sizable peak in the number of reassessments of the disability grade at the end of the first stage.

recovery from sickness and the institutional incentives for work resumption. The lowest wage group shows a dramatic drop at around 60 months. This might owe to the small number of individuals with a long duration of DI receipt in the study sample (see Figure 4).

The relationship between labor participation and age is less consistent apart from a generally strong decrease for individuals older than 55 years. It might be that job opportunities are limited for workers who are older and partially disabled at the same time.

Labor participation shows a strong positive relationship with calendar time. This relationship is potentially driven by the duration of DI receipt. For example, duration of DI receipt is low when individuals enter the DI scheme where at the same time they have low levels of labor participation. Consequently, in the beginning of the observation period, corresponding to the period individuals enter the DI scheme, labor participation is low. The somewhat smaller increase starting in the beginning of 2011 is likely to be driven by the lack of entry to the DI scheme from this point onwards since sick individuals tend to have low labor participation when they enter the DI scheme. Age, and possibly also the 2007-2008 financial crisis, could contribute to the calendar time profile of labor participation.

Duration of DI receipt, age and calendar time are, for every individual, perfectly correlated with event time. Event time, as we define it in Section 5.3, indicates the period of time before individuals face the work resumption program, and subsequently the period of time when individuals face the program. Since the treatment status of having faced the program depends on event time by definition, not accounting for the effects of these three variables could bias the treatment effect estimate. The next section elaborates on this.

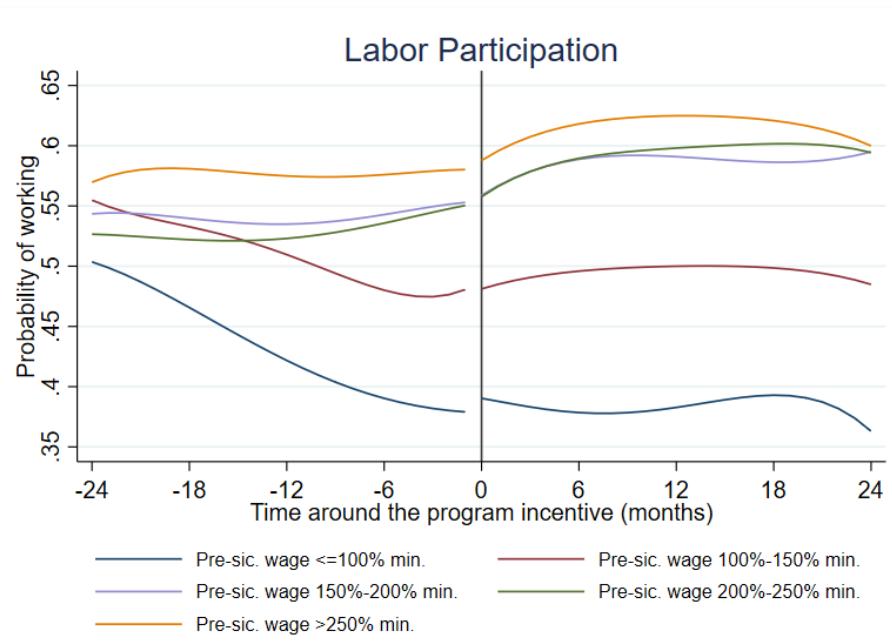


Figure 3: Labor participation around the time the work resumption program takes effect.

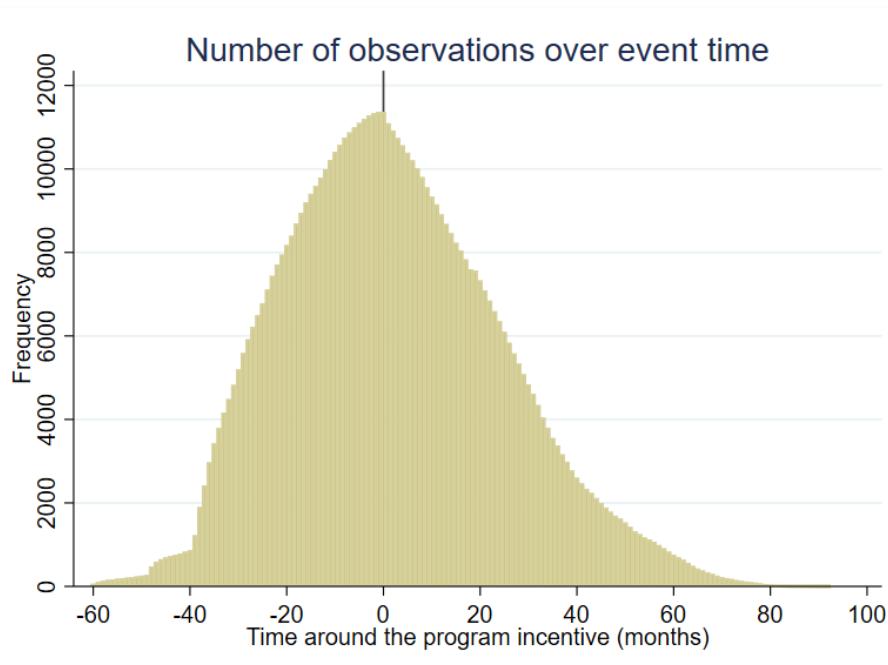


Figure 4: Distribution of the duration of DI receipt around the time the work resumption program takes effect.

Table 1: Sample means of background characteristics and labor market outcomes of select pre-sickness wage groups in the month before and the month after the work resumption program takes effect

	Pre-sickness wage less than or 100% of minimum wage			Pre-sickness wage 150–200% of minimum wage			Pre-sickness wage >250% of minimum wage		
	Before	After	Diff.	Before	After	Diff.	Before	After	Diff.
Age	46.37	46.46	0.08	48.47	48.55	0.08	53.42	53.50	0.08
Male	0.17			0.52			0.75		
Married	0.54	0.54	0.00	0.51	0.51	0.00	0.63	0.63	0.00
Disability grade >50%	0.26	0.25	-0.01	0.50	0.51	0.00	0.45	0.45	0.00
Months receiving wage-related benefit	23.35			26.28			30.88		
Months receiving wage-supplement benefit	10.74			16.32			15.01		
Months receiving follow-up benefit	19.04			13.02			11.19		
Pre-sickness daily wage	48.44	48.42	-0.02	103.04	103.07	0.03	181.93	181.96	0.03
With social minimum supplement	0.25	0.29	0.04	0.00	0.11	0.11***	0.00	0.07	0.07***
Labor participation	37.98	38.81	0.83	55.45	56.10	0.65	58.14	58.56	0.42
Daily wage while disabled	9.94	10.20	0.26	28.50	28.83	0.33	48.96	49.44	0.48
Sample size	603	603		3,212	3,212		2,377	2,377	

Notes: Labor participation is defined as having positive wage earnings. Pre-sickness daily wage and daily wage while disabled are in January 2006 euros (derived). ***, **, * denote statistical significance at 1%, 5% and 10%, respectively.

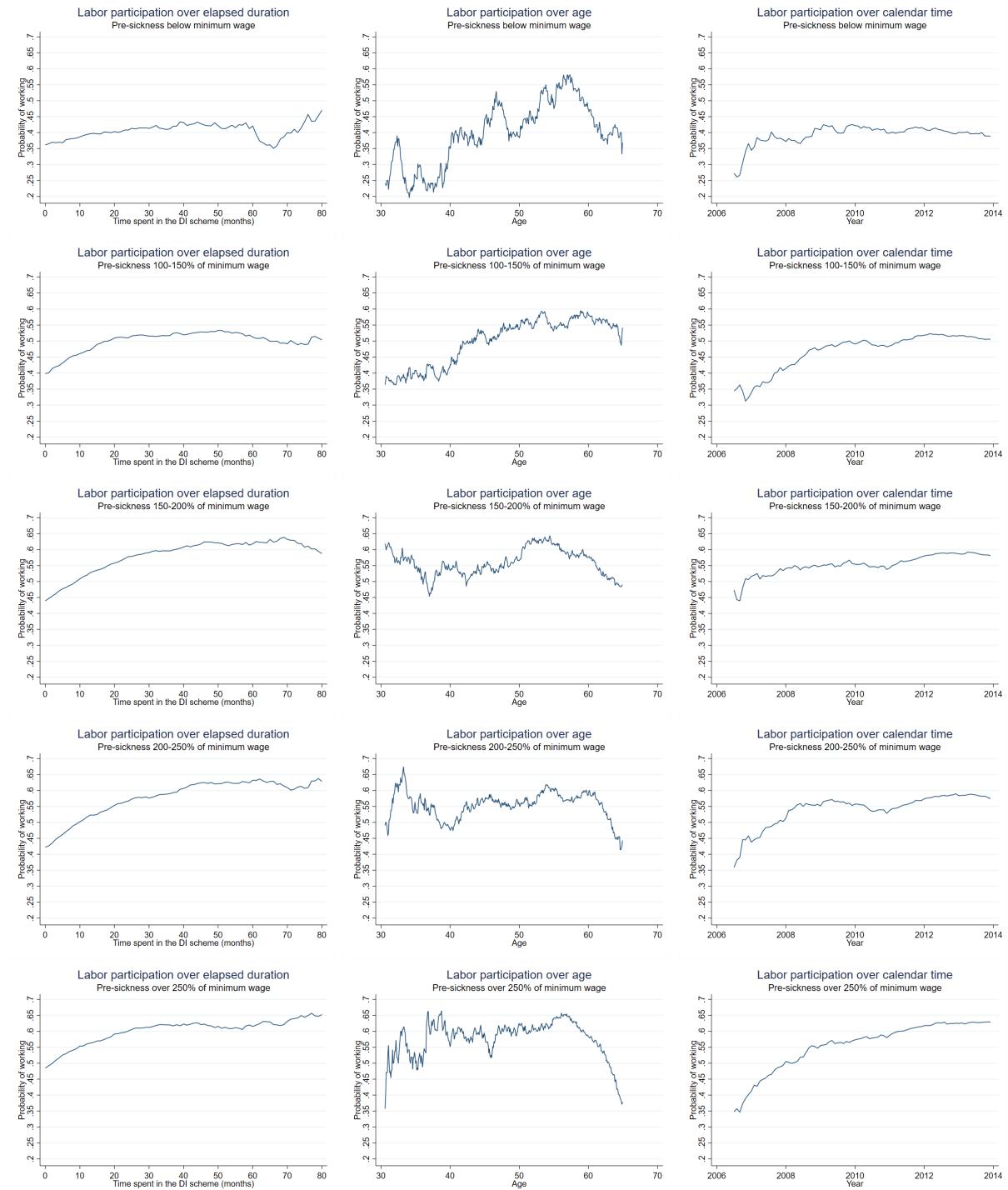


Figure 5: Labor participation against duration of DI receipt, age and calendar time by pre-sickness wage groups. Regions with small sample sizes are not shown because of high volatility.

5 Identification strategy

As described in Section 2, in the work resumption program individuals with higher pre-sickness wages face a stronger incentive to work when their first-stage DI benefit expires. Therefore, we distinguish across five pre-sickness wage groups to analyze if the program leads to higher labor participation responses among those with higher pre-sickness wages.

The estimation strategy to identify the effects of the financial incentives is based on a Regression Discontinuity (RD) design. This design requires a cutoff in the running variable where the probability of treatment changes discontinuously. Our running variable is the time around the moment individuals move from the first to the second stage of the DI scheme. It is defined such that it equals 0 in the first (event) month the individual receives a second-stage benefit (wage-supplement or follow-up), 1 in the month after, -1 in the month before, etc. By this construction, probability of treatment is 0 for all negative values of the running variable, and 1 for all nonnegative values. Since the running variable takes on only integer values, the cutoff can in principle be placed anywhere between -1 and 0 without affecting the treatment assignment for any of the observations. This gives rise to a sharp RD design as compliance is perfect.

As the running variable takes on discrete values, use of the standard continuity-based RD approach is questionable, especially given the fact that the number of distinct values of the running variable is as small as 153. That is, when the running variable is discrete and a continuity-based RD design is applied, the effective number of observations used is the number of the distinct values, not the total number of observations (Cattaneo et al., 2020). For our study, this means estimating the treatment effect using 153 distinct values of the running variable, instead of the available 633,251 monthly observations for 11,366 individuals. In this case Cattaneo et al. explain that the Local Randomization (LR) approach to RD may be the only valid RD approach. This approach requires stronger assumptions than the continuity-based RD approach. However, if its assumptions are met, it allows the running variable to be discrete and to identify still a causal effect.

The rationale is that there exists a window around the cutoff where the data behaves as if it were part of an experimental setup in which units receive a random score (value of the running variable), and are assigned to treatment if and only if their score is equal to or above the cutoff value. Consequently, the value of the running variable is not related to any observed or unobserved characteristic of the units, other than by chance. As long as the assigned score by itself does not affect the outcome, the expected value of the outcome variable, conditional on treatment status, is then the same for all values of the running variable within the window, even if the control and treatment groups are only observed in (mutually exclusive) parts of the window. This is illustrated in Figure 6. The difference between the expected values of the control and treatment groups identifies the treatment effect. This difference can be estimated by the difference in the sample averages of the two groups.

The LR approach to RD identifies the treatment effect around the cutoff. As we do not take a continuity-based approach to RD, we do not make continuity assumptions to identify and estimate a causal effect at the cutoff (Mattei and Mealli, 2017). Furthermore, Cattaneo et al. (2015), Mattei and Mealli and Sales and Hansen (2020) propose different sets of assumptions within a neighborhood of the cutoff to interpret a RD design as a local randomized experiment. We adapt the approach of Cattaneo et al..

5.1 Identification assumptions

Let r_0 be the cutoff value of the running variable, T the treatment indicator, and W_0 the local randomization window. The potential outcomes for hypothetical values of the running variable

and treatment status of individual i are denoted with $Y_i(r, t)$. There are two assumptions that need to hold for the validity of the LR approach (henceforth LR assumptions).

LR Assumption 1: Unconfounded assignment. The distribution function of the running variable inside the window, $F_{R_i|R_i \in W_0}(r)$, does not depend on the potential outcomes and is the same for all units: $F_{R_i|R_i \in W_0}(r) = F_0(r)$, where F_0 is any distribution function.¹⁰

LR Assumption 2: Exclusion restriction. The potential outcomes do not depend on the value of the running variable inside the window, except via the treatment assignment indicator: $y_i(r, t) = y_i(t) \forall i$ such that $R_i \in W_0$.

The first identifying assumption states that for all units within the window, the distribution of the running variable is the same. In our analysis the running variable is determined as follows. For each individual it starts at a certain negative value, which depends on the work history, and it incrementally increases over time until the individual exits the scheme or until the end of the observation period.¹¹ Therefore, individuals cannot manipulate the value of the running variable when they are in the DI scheme. Furthermore, due to the panel nature of the available data, we have observations on (mostly) the same people before and after the cutoff. Therefore, people on both sides of the cutoff should have similar (unobserved) characteristics. These lend strong support to that the distribution of the running variable is the same for all units.

There is, however, a caveat regarding the timing of entry into and exit from the DI scheme. If this was the same for all individuals, they would all have experienced same values of the running variable, regardless of their observed (e.g. duration of the first-stage benefit) or unobserved characteristics. In particular, this would imply that their potential outcomes cannot affect the running variable. In fact, this holds for the smallest possible window, $\{-1, 0\}$. Everyone in the study sample experiences both the last month of the first stage and the first month of the second stage of the DI scheme. However, as the window widens, more and more individuals have missing observations within the window at one or both sides of the cutoff. These missing observations may not be distributed randomly across individuals. The duration of the first stage of the DI scheme, and thereby the number of observations below the cutoff, is determined by the work histories of individuals. Individuals with long work histories may systematically differ from those with short work histories. For example, individuals with higher potential outcomes may have been more likely to work before falling sick and therefore have a longer first stage duration. Similarly, people who exit the DI scheme soon after the cutoff may be systematically different from those who do not, although a relatively small part of the sample (14.5%) exits the scheme before the end of the observation period, of whom less than half exit in the first 12 months after the cutoff. To address possible effects that could confound the treatment effect due to non-random missing data, in a sensitivity check we compare results from the full sample with results from a restricted sample including only the individuals who experience the entire window.

The second identifying assumption states that, within the window, the potential outcomes do not depend on the running variable except through the treatment. In other words, if there were no difference in treatment status between observations, the expected value of the outcome should be the same for all values of the running variable within the window W_0 . Indeed, due to the exogenous variation in the duration of the first stage, and hence the moment of entry

¹⁰We do not assume that the distribution function is known. This is not needed because in the analysis we do not make use of finite-sample inference due to the large sample available (Cattaneo et al., 2020).

¹¹Duration of the first-stage benefit is determined by the duration of unemployment insurance which depends on the work history (Section 2).

into the DI scheme in terms of calendar time, the month of the cutoff is like any other month, and therefore it is not immediately clear why being in a certain month before or after it should matter for the outcome (labor participation) in the absence of treatment.

However, the running variable does contain information about the passage of time. For each individual this variable is perfectly correlated with other variables related to time such as time spent in the DI scheme, age, and calendar time. As shown in Section 4, these variables may affect labor participation. Below we discuss how we account for these variables in the regression analysis.

Exploiting the panel dimension of the available data, we control for individual-specific fixed effects to account for time-invariant observed and unobserved characteristics that may affect labor participation. Consequently, the second identifying assumption implies the following. Conditional on time-variant controls and time-invariant individual attributes, the running variable does not affect the potential outcomes within the window other than through treatment assignment.

To identify the full treatment effect, we require additional assumptions related to the timing of the responses to the financial incentives. Since the model is based on a comparison of the outcome before and after the cutoff, we need that people close to the cutoff, but still in the first stage of the scheme, do not already change their labor participation in anticipation of the program incentives of the second stage. The modest increase in labor participation of all but the lowest wage group in Figure 3 might cast doubt on this assumption, as it could be explained by anticipation that is stronger for higher pre-sickness wage groups.

There can be, however, alternative explanations for the different trends observed just before the cutoff. First, changes in the outcome prior to the treatment could be due to that treatment is endogenous, instead of that they are due to anticipation (Malani and Reif, 2015). This is, however, impossible since the duration of the first stage of the DI scheme is institutionally determined and cannot be adapted in response to the changes in labor supply, or to factors correlated with the labor supply of beneficiaries. Second, what looks like an anticipation effect in higher pre-sickness wage groups can be a sample composition effect. In general individuals with short work histories necessarily make up a larger fraction of a given sample before but close the cutoff (see Figure 4). Higher pre-sickness wage groups before but close to the cutoff may include fewer individuals with short work histories who at the same time have low labor participation rates when they enter the DI scheme. In fact, in the study sample, the fraction of people with a first-stage duration of at most 3 months is 0.9% in the highest wage group while this is 3.0% in the lowest wage group. Furthermore, considering people in all wage groups with a first-stage duration of at most 3 months, the fraction of those working is 19.0% in the highest wage group while this is 11.1% in the lowest wage group.

Still, for higher pre-sickness wage groups, if part of the labor supply response already takes place before the cutoff, the treatment effect will be underestimated. A similar issue arises for the labor supply response after the cutoff. If individuals are able to utilize their remaining work capacity in suitable jobs at any given point in time, they could react to the program incentives immediately after the cutoff. In practice, however, it may be unrealistic to expect individuals to time their response so precisely. The labor supply response may be spread out farther beyond the cutoff due to adaptation. In this case the RD estimator will, again, underestimate the true treatment effect.

Underestimation of the treatment effect due to anticipation before the cutoff, or adaptation after the cutoff, is not solved by taking a wider window, although the estimates may be closer to the true effect. For the latter, for example, consider the case where the no-anticipation assumption holds, but after the cutoff it takes x months ($x > 0$) before all program participants have responded. Conditional on covariates and individual fixed effects, the estimation in the

window $\{-x - 1, \dots, x\}$ compares the mean in the $\{0, \dots, x\}$ window to the mean in the $\{-x - 1, \dots, -1\}$ window. However, the average effect in the $\{0, \dots, x\}$ window will be below the full effect at x .

These imply that the estimated effects of treatment can be viewed as lower bounds to the true full treatment effects. However, even if the assumptions regarding the timing of the response are violated, the differences between pre-sickness wage groups remain equally informative as long as individuals in the different groups violate the assumptions in the same way.

5.2 Window selection

An important step in the LR design is window selection. Ideally, the selected window is the widest possible window within which the LR assumptions hold. However, this cannot be tested directly. Cattaneo et al. (2020) describe two options to select the window. The first option takes a data-driven approach. The window size estimate is the window within which one or more observed covariates do not change with the running variable but can change outside it. Such a covariate does not exist in our data. This leaves us with the second option, which is choosing the window in an ad-hoc manner. We consider a window size that is symmetric around a hypothetical value of -0.5 , and therefore contain the same number of months before and after the cutoff.

Several considerations should be weighed against each other when choosing the window. On the one hand, the wider the window, the more precisely coefficients can be estimated, as more observations are used. On the other hand, the smaller the window, the more likely that the LR assumptions hold. As (symmetric) windows are nested within each other, it follows that if there is some window W_0 in which the LR assumptions hold, they will necessarily hold in smaller windows. However, as the window gets wider, the less plausible it becomes that there is no relation between the running variable and the outcome variable anywhere in the window. In the context of the current study, the further away in time from the cutoff, the more likely that there are changes in (unobserved) individual characteristics that may affect labor participation, such as health or household income. If these changes generally move in a specific direction over time, for example health worsens over time, estimation of the treatment effect will be biased.

In terms of the assumptions related to the timing of the response (no anticipation before the cutoff and no adaptation after it), the size of the window seems irrelevant. The assumption that the effect of treatment is observed immediately after the cutoff needs to hold in every window, because every window includes the immediate term. Furthermore, it is hard to conceive of a situation in which there is an anticipation effect some time before the cutoff but not right before it. Only then would the assumption of no anticipation hold in small windows but not in wider ones.

Given the large sample sizes in all pre-sickness wage groups, the variance of the coefficient estimates is less of a concern than the potential bias introduced if the LR assumptions do not hold. Even in the smallest possible window there are over 1,200 observations in the data for each group. Therefore, the smallest possible window is preferred in this study, which is $\{-1, 0\}$. We consider alternative window sizes to show sensitivity.

5.3 Regression specification

We estimate the following regression model:

$$y_{ir} = \alpha_i + \beta T_{ir} + X_{ir}\gamma + \varepsilon_{ir}. \quad (1)$$

i indexes individuals. r indexes the event time. Its values from -60 to -1 indicate the months before individuals face the work resumption program, 0 is the cutoff month when first subjected

to the program, and 1 to 92 are the remaining months of the program. The dependent variable, labor participation, is given by y . The coefficient β on the treatment indicator T is the main parameter of interest. Under the identifying assumptions, it captures the mean effect of the work resumption program around the cutoff. X is the vector of covariates related to the running variable. It includes dummies controlling for multiples of six months of DI receipt duration, and a linear function of the business cycle indicator. The dummies capture the strong relationship between labor participation and DI duration as suggested by Figure 5. The business cycle indicator captures macroeconomic shocks. We consider a linear functional form as suggested by the binned scatter plot of labor participation against the (lagged) indicator in Figure 18 in the Appendix.^{12,13} α_i is an individual-specific constant. It captures time-invariant labor participation differences across individuals. It also captures differences in pre-sickness wage levels or time-invariant health conditions. ε_{ir} is an idiosyncratic (unobserved) shock, assumed to be uncorrelated with all explanatory variables. Only observations for which the value of the running variable falls within the window, $r \in W_0$, are used in the estimation.

¹²The indicator is constructed by the Dutch Central Bank to identify turning points in the Dutch business cycle (Butler et al., 2019). It is composed of 86 potential sub-series that are closely related to the development of real GDP growth and are also up to six months ahead of it. Figure 17 in the Appendix shows that the indicator is reasonably capable of identifying tipping points in real GDP growth from the recent past. We allow a period of 6 months for the indicator to affect labor participation. Since the indicator itself predicts 6 months ahead, we lag the indicator by 12 months. When the indicator is lagged 12 months, Figure 18 shows a positive impact on labor participation as we would expect. It shows a negative impact when lagged 6 months.

¹³We cannot consider calendar time and age in the vector of covariates because they are perfectly collinear with the dummies for DI duration. We consider that the business cycle indicator captures the macroeconomic shocks that calendar time would capture. To investigate the possible effect of age on the treatment, we conduct heterogeneity analysis with respect to age.

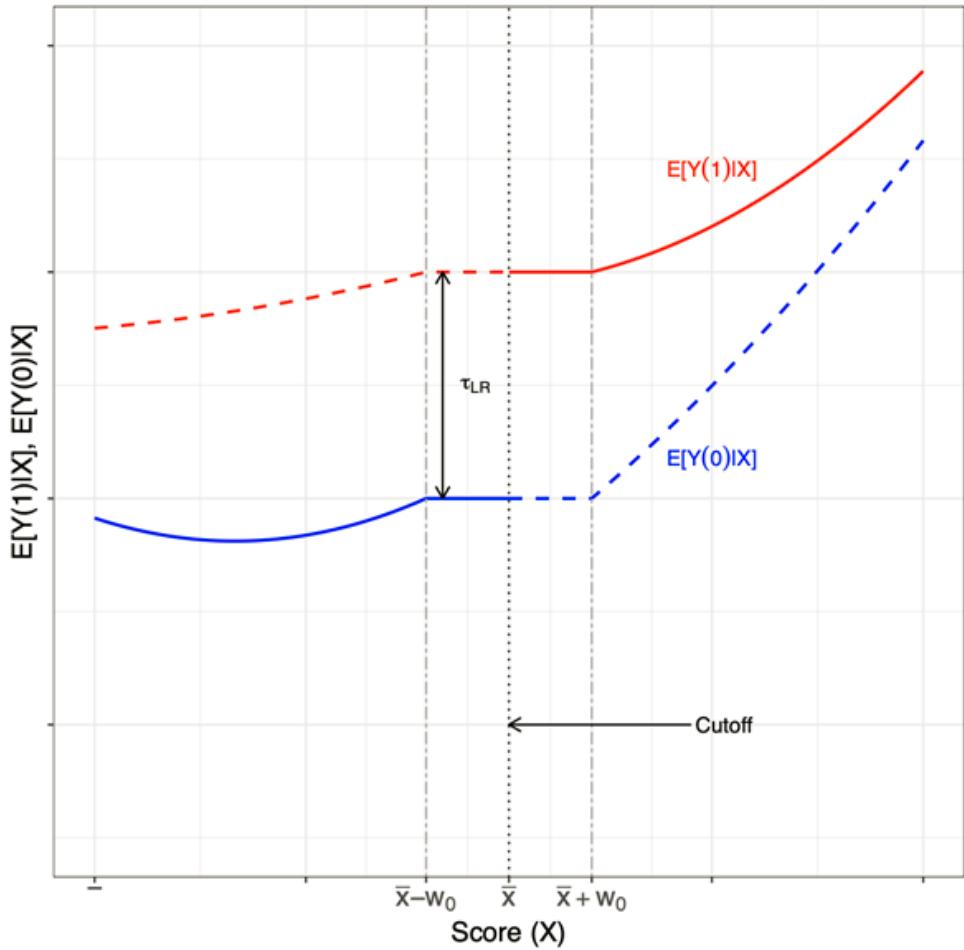


Figure 6: Illustration of the local randomization approach to RD. Source: [Cattaneo et al. \(2020\)](#).

6 The effect of the work resumption program on labor participation across pre-sickness wage groups

Table 2 presents the estimated treatment effects for five income groups in the $\{-1, 0\}$ window.¹⁴ The estimation is based on the baseline specification defined in Section 5.3. For individuals with pre-sickness wages below the minimum wage or 100-150% of the minimum wage, the work resumption program has no statistically significant effect on labor participation. On the other hand, for individuals with pre-sickness wages 150-200% or 200-250% of the minimum wage, the effect is significant. For these income groups labor participation increases by 0.6 and 0.5 pp, respectively. For the highest income group the effect is insignificant.

These results are in line with the program incentives which are stronger for individuals with higher pre-sickness wages (Figure 2a). They are also in line with the labor participation trends in Figure 3. The figure shows that, relative to the first stage of the DI scheme, labor participation probabilities of the three highest income groups increase in the second stage when financial incentives to resume working take effect. Lower income groups, however, are not able to increase labor participation in the second stage, for two potential reasons. First, they struggle to cope with their disability and increase labor supply. Second, they face smaller financial incentives to resume working.

The point estimates in Table 2 are based on a window size of two months and show the immediate effects of the work resumption program. However, if some individuals change their labor participation already before the cutoff in anticipation of the financial incentives, or if they adapt to the incentives only some time after the cutoff, the estimates based on the smallest window are an underestimation of the full treatment effect of the financial incentives. If this is the case, estimation of the treatment effect from wider windows around the cutoff should lead to larger estimates. To test this, we estimate the treatment effect from wider windows. However, as argued in Section 5, if the assumptions related to the timing of the response do not hold, estimations based on wider windows do not identify the full treatment effect, but they may approximate it. Furthermore, results from wider windows should be interpreted with caution as the LR assumptions may be violated in these windows.

For the five pre-sickness wage groups, Figure 7 presents point estimates of the program effect based on symmetric windows around the cutoff of smallest size of 2 months up to 48 months, corresponding to windows $\{-1, 0\}$ up to $\{-24, \dots, 23\}$, respectively. There are two main findings. First, the treatment effect estimate is larger in wider windows than in the smallest window, especially at windows of sizes up to 8, possibly due to anticipation and adaptation effects as suggested above. In Figure 8 we investigate the probability of working conditional on not working in the previous month. There is a global trend of decreasing work resumption throughout the window of 48 months around the cutoff. A potential reason is that partially disabled individuals have stronger labor market attachment and hence work resumption early on during their disability than later. The more variable work resumption around 20 months before and after the cutoff can be explained by smaller numbers of individuals with long first- and second-stage durations (Figure 4). The decreasing trend, however, is interrupted by increased work resumption in the two months preceding and in the three months succeeding the cutoff, providing supportive evidence that individuals increase their labor participation already before the cutoff in anticipation of the program, and respond with a delay due to adaptation after they face the program the first time at the cutoff.

Second, the effects are larger for higher wage groups. This finding is in line with the

¹⁴In the regressions presented in this section, the dummies controlling for multiples of six months of DI receipt duration are almost always significant except in the smallest window and for the lowest wage group. The linear function of the business cycle indicator is almost always insignificant.

heterogenous incentives of the program which are stronger for higher wage groups. There is limited evidence of a program effect among individuals with pre-sickness wages below the minimum wage. The estimates are significant when window sizes are 4, 6 and 8 months, but they are smaller and turn insignificant for wider windows. The effects are somewhat stronger for individuals with pre-sickness wages 100-150% of the minimum wage. For higher wage groups, the patterns of the estimates are qualitatively similar to each other, especially those of the estimates based on windows of sizes up to 24 months. For the three smallest windows, the estimated treatment effect increases notably, with effect sizes of about 0.5 up to 1.5 pp. For the subsequent three wider windows, the impact of the program increases further. This suggests that a large part of the effect translates into labor participation within three months after the cutoff. For subsequent wider windows the patterns stabilize for all three groups. From a window size of 24 months, low wage groups do not increase labor participation while higher wage groups increase it considerably. The increases are larger for higher wage groups. Again, this finding is in line with the heterogenous incentives of the program.

Considering a window size of 6 months, the estimate of the program effect is 1.1 pp on average over all wage groups. This is close to the average short-term effect estimated by [Koning and van Sonsbeek \(2017\)](#) using the full sample that is of size 1.4 pp, despite that their estimation method is different. In particular, they compare the 6 months around the cutoff to the period before, instead of how we compare 3 months after the cutoff to 3 months before. On the other hand, considering the widest window around the cutoff, the estimate of the program effect is 2.0 pp on average over all wage groups. This is somewhat smaller than the long-term effect estimated as 2.6 pp by [Koning and van Sonsbeek](#) based on a comparison of the period more than 3 months after the cutoff to the period more than 3 months before the cutoff.

The long-term effect of the program we estimate should be interpreted with caution, however. In the LR design, potential outcomes can depend on the running variable outside the window W_0 (Figure 6). This is because the LR assumptions do not need to hold outside the window. This, however, also means that the LR estimates outside the window do not need to reflect causal effects. Therefore, we take the long-term effect estimates, that is those from a window size of 24 months, as suggestive of causal effects.

Table 2: Impact of the work resumption program on labor participation among pre-sickness wage groups

	Labor participation
Pre-sickness wage <100% of min. wage	0.012 (0.008)
Pre-sickness wage 100–150% of min. wage	0.000 (0.003)
Pre-sickness wage 150–200% of min. wage	0.006** (0.003)
Pre-sickness wage 200–250% of min. wage	0.005* (0.003)
Pre-sickness wage >250% of min. wage	0.006 (0.004)

Notes: The window size is 2 months in each regression. The regressions are based on 1,206, 5,192, 6,424, 5,156 and 4,754 observations, respectively for the lowest and higher wage groups. Regressions control for covariates and fixed effects. ***, **, * denote statistical significance at 1%, 5% and 10%, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level.

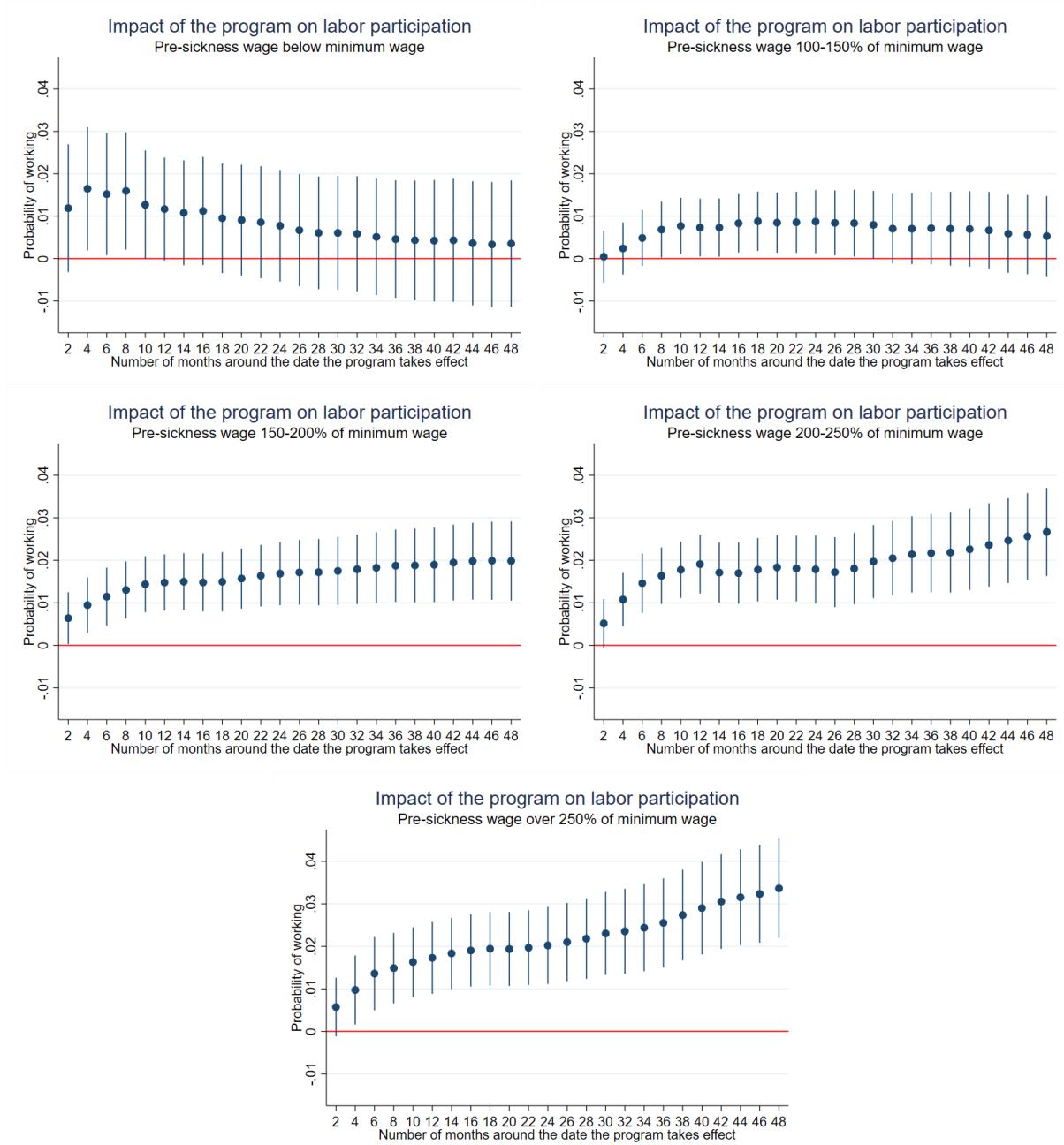


Figure 7: Treatment effect estimates based on the baseline specification for different window sizes. Vertical lines indicate the 95% confidence intervals.



Figure 8: Probability of working conditional on not working in the previous month during a period of 48 months around the cutoff.

7 Falsification tests and sensitivity analyses

7.1 Window selection and identification

The exclusion restriction assumption of the local randomization method requires that the potential outcomes do not depend on the value of the running variable inside a window, except via the treatment assignment indicator (Figure 6). The baseline results in Figure 7 show that the treatment effect estimate, as the estimate of the difference in the expected values of the potential outcomes, show an increase, mostly for higher wage groups. As discussed in Section 5.1, this can be due to, for example, that individuals are not able to react to the program incentives immediately after the cutoff, especially those in higher pre-sickness wage groups who face larger incentives and expected to react. Besides this increase, however, the treatment effect estimate is fairly stable except for largest window sizes, suggesting that the potential outcomes do not depend on the value of the running variable, except through treatment assignment. Here we check to what extent the treatment effect estimate remains stable across wider window sizes in a statistical sense.

We consider the p-value from a two-sample z-test of the difference between the treatment effect estimate from a window size and that from a benchmark window size of 10 months. The benchmark is determined based on the observation that, in Figure 7, treatment effect estimates start to stabilize from the window size of 10 months for all wage groups.

Figure 9 presents the p-value for all window sizes except for the window size of 10 months as the benchmark for comparison. With few exceptions, the p-values are far larger than the conventional p-values for statistical significance. These results suggest that, in a statistical sense, the treatment effect estimate is stable across a wide range of windows around the cutoff providing evidence for the identifying exclusion restriction assumption.

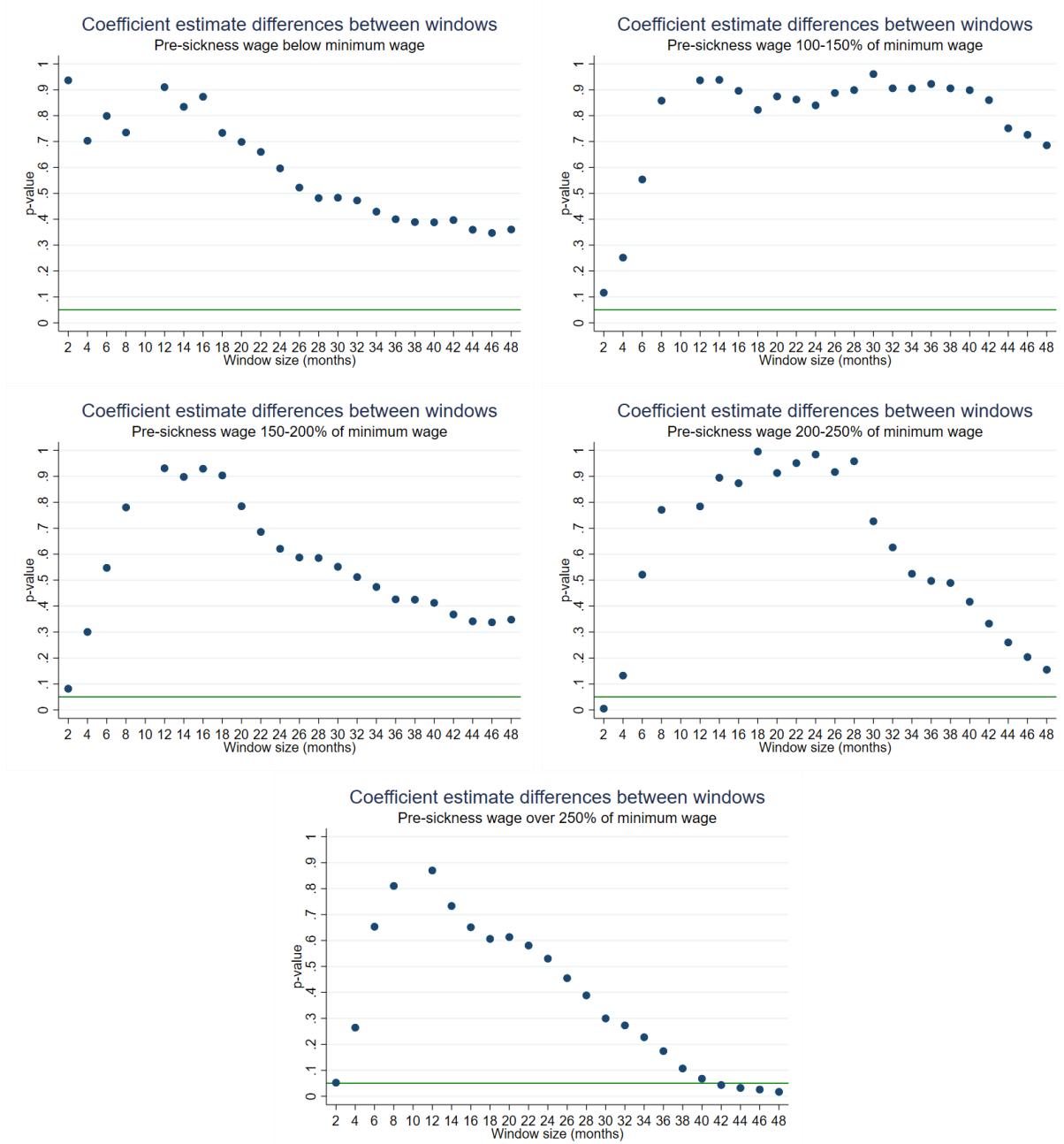


Figure 9: p-values of the test of the difference in treatment effect estimates from different window sizes and that from a window size of 10 months. Green horizontal line marks the p-value at 5%.

7.2 Donut hole regression

As discussed in Section 5.1, to identify the full treatment effect, we require that individuals do not change their labor participation already before the cutoff in anticipation of the work resumption program, or do not respond with a delay due to adaptation after they face the program the first time at the cutoff. Assuming that anticipation and adaptation are part of the true responses, we investigate to which extent they lead to an underestimation of the full treatment effect in the baseline regression.

We estimate a “donut hole” regression where we exclude the data in the window of 8 months, keeping otherwise the baseline regression specification the same. A window of 8 months is chosen based on the observation that, in Figure 7, treatment effect estimates show a notable increase in windows of 2 to 8 months in almost all wage groups, possibly due to anticipation or adaptation effects.

The center panel of Figure 10 presents the estimation results based on the donut hole regression. There are two main findings. First, except for the lowest wage group, the treatment effect estimates are larger than those based on the baseline regression shown in the left panel. As Figure 8 suggested, this is potentially due to anticipation and adaptation effects. Second, in line with the baseline results, the results based on donut hole regressions provide clear evidence of heterogeneous treatment effects across income groups. The highest wage group increases labor participation by 4 to 5 pp, the second highest by about 4 pp, the third by about 3 pp, the second by 1 to 2 pp, and for the lowest wage group the effect is statistically not different from 0.

In the right panel of Figure 10 we investigate whether the difference between the baseline and donut hole regression estimates of the treatment effect is statistically significant. We consider the p-value from a two-sample z-test of the difference between the two types of estimates. For the three highest wage groups, the difference between the two estimates are statistically significant at conventional levels for most window sizes. This implies that, for these cases, baseline treatment effects might be underestimated due to anticipation of the program or adaptation to it. In these cases the baseline treatment effect estimates can be viewed as lower bounds to the full treatment effects.

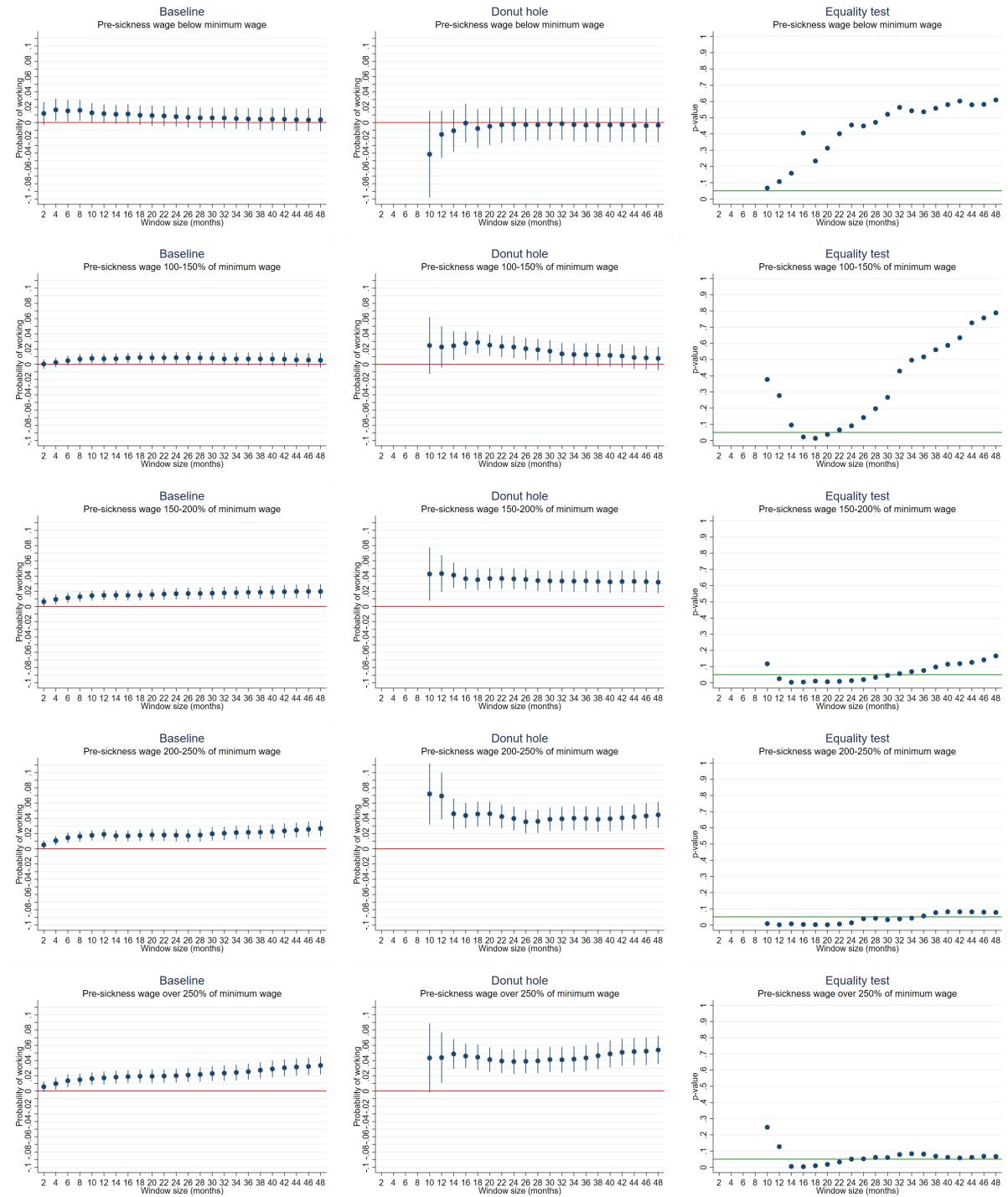


Figure 10: Treatment effect estimates from the baseline (left panel, reproducing Figure 7 with y-axis rescaled) and donut hole (center panel) regressions, and p-values of the test of the difference in these estimates (right panel). Vertical lines indicate the 95% confidence intervals. Green horizontal line marks the p-value at 5%.

7.3 Placebo cutoffs

The baseline results have shown that there are significant differences around the cutoff, comparing individuals before and after they face the work resumption program. In order to assess that these differences are caused by the treatment and not by other factors, we compare the mean outcome around placebo cutoffs, before and after the true cutoff, keeping otherwise the regression specification the same as in the baseline. At placebo cutoffs before the true cutoff, this leads to a comparison that is between individuals who have not yet been treated. After the true cutoff, the comparisons are between treated individuals. Significant differences around placebo cutoffs may cast doubt on that the differences around the true cutoff are due to the treatment. In the context of the current study, however, it should be noted that the results at placebo cutoffs close to the true cutoff may be indicative of an anticipation or adaption effect in response to the program incentives.

Table 3 presents results based on placebo cutoffs 3 and 12 months before and after the true cutoff, based on a window size of 2 months. There are a number of significant differences between the placebo treatment and control groups. These significant differences seem closely related to the work resumption trends observed in Figure 8. Placebo effects visible at 3 months before and after the true cutoff are potentially due to that individuals respond to the program around these cutoffs due to anticipation and adaptation. The effects visible at 12 months before the true cutoff but not at 12 months after it are potentially due to that work resumption is more likely early on during DI participation than later. Therefore, these results do not cast doubt on that the differences around the true cutoff are due to the work resumption program.

Table 3: Labor participation around placebo cutoffs

	Placebo cutoff 12 months before the true cutoff	Placebo cutoff 3 months before the true cutoff	Placebo cutoff 3 months after the true cutoff	Placebo cutoff 12 months after the true cutoff
Pre-sickness wage <100% of min. wage	0.000 (0.006)	0.006* (0.004)	-0.000 (0.003)	0.005 (0.003)
Pre-sickness wage 100–150% of min. wage	0.007*** (0.002)	0.006*** (0.002)	0.004 (0.003)	0.004* (0.002)
Pre-sickness wage 150–200% of min. wage	0.008*** (0.002)	0.000 (0.002)	0.005** (0.002)	-0.001 (0.002)
Pre-sickness wage 200–250% of min. wage	0.005** (0.002)	0.005** (0.002)	0.007** (0.003)	0.001 (0.002)
Pre-sickness wage >250% of min. wage	0.000 (0.002)	-0.003 (0.002)	0.005** (0.002)	0.002 (0.001)

Notes: The window size is 2 months in each regression. The number of months before or after is in relation to the date of the true cutoff. ***, **, * denote statistical significance at 1%, 5% and 10%, respectively. Standard errors (in parentheses) account for heteroskedasticity and clustering at the individual level.

7.4 Covariate balance

If the local randomization assumptions hold, the treatment should not have an effect on any covariate whose values are realized before the treatment is assigned. To test this assumption we use regression where the dependent is a covariate and the predictor is the treatment indicator. A significant treatment indicator suggests that treatment assignment is not random due to the covariate. In this regression we allow for individual fixed-effects as we allow for them in the baseline regression. It is not possible to test whether unobserved time-varying characteristics are balanced around the cutoff. However, if the observed covariates are balanced, it is plausible that the unobserved covariates are balanced as well.

Note that, in our setting, it should be unlikely that covariates are imbalanced due to treatment since individuals face the treatment at random times depending on the number of years they work before falling sick. Hence, any observed imbalance is likely be due to factors other than treatment. We discuss alternative factors below.

The left panel of Figure 11 shows, across pre-sickness wage groups and alternative window sizes, covariate balance for disability grade being higher than 50%.¹⁵ It is balanced for all wage groups in the smallest window. However, it shows an increasing disparity through wider windows, except for the lowest wage group. This suggests that health deteriorates over time. An alternative explanation is that over time individuals recover and leave the DI scheme, and hence the sample data, and therefore the sample consists increasingly more people with higher disability grades at wider windows.

Still, when wider windows are considered, our estimates of the treatment effect could be confounded due to a health effect. However, this is unlikely. The lowest wage group does not show a statistically significant health effect of treatment. The baseline estimates showed systematic differences of the treatment effect on labor participation across the wage groups. If these differences were due to health differences across the wage groups, we should observe deteriorating health more often for lower than higher wage groups. However, except the lowest wage group, higher wage groups share a very similar pattern of the change in disability grade across wider windows. Furthermore, if our estimate of the treatment effect is reflecting the effect of deteriorating health, it will be biased downward instead of upward.¹⁶

The middle and right panels of Figure 11 show regression results where dependent is gender or duration of the first-stage of the DI scheme, and the predictor is the treatment indicator where we do not control for individual fixed effects. Both outcomes are predetermined and time-invariant for each individual. However, in most cases, within wage groups, their sample average changes over time. For wide windows, the difference in means before and after the cutoff is sometimes substantial. This is due to entry in and exit from the data, the timing of which is thus not equally distributed across men and women, and across people with different work histories (which determine the duration of the first stage).¹⁷ As we control for such time-invariant covariates in regression analyses, this should not compromise the identifying assumption of unconfounded assignment.

¹⁵It is possible that disability grade is not entirely exogenous to the treatment. Due to capacity constraints, only a small fraction of individuals are reassessed of their disability grade in a given period by the Employee Insurance Agency. If wage earnings in relation to the assessed remaining earning capacity is a criterion to select someone for reassessment, treatment might affect this through labor supply decisions. Still, it is not obvious whether a causal relation exists in the direction from treatment to disability category.

¹⁶We also checked and found that indicator of being married or in a registered partnership as a time variant covariate is balanced around the cutoff for all wage groups and window sizes.

¹⁷Exit from the data does not necessarily mean exit from the DI scheme. Most individuals are observed until the end of the observation period.

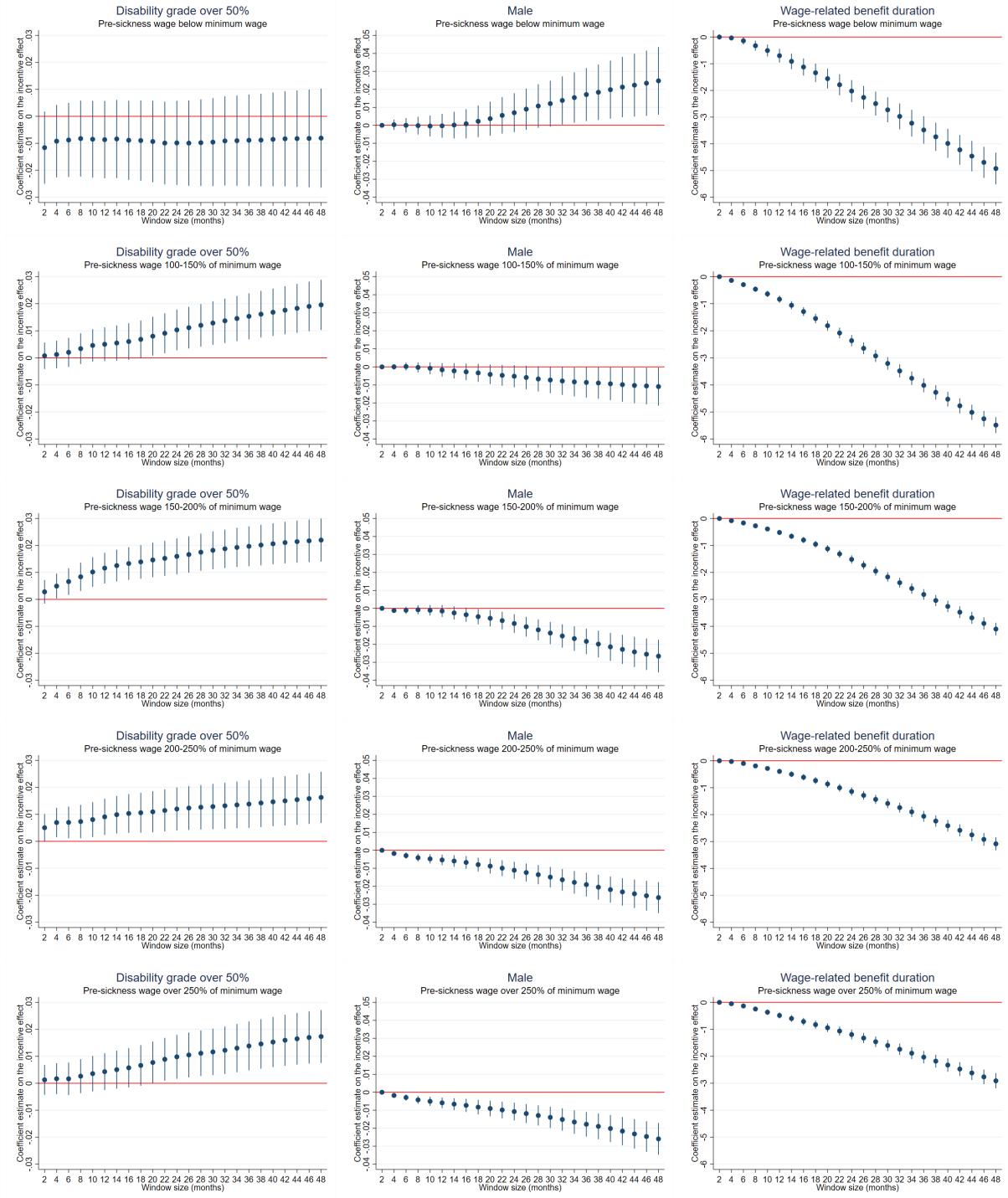


Figure 11: Effect of the treatment on observable characteristics, without controlling for individual fixed effects. Vertical lines indicate the 95% confidence intervals.

7.5 Changes in the sample composition

The identifying assumption of unconfounded assignment requires that every individual is equally likely to be observed anywhere within the window where local randomization occurs (Section 5). The sample composition may change over the course of event time (the running variable) due to entry into and exit from the DI scheme. If timing of entry and exit is random across individuals, compositional changes are inconsequential for the identifying assumption. However, if it is not random and related to certain individual characteristics, these individuals will not be equally represented across the values of running variable and violate the identifying assumption.

In Figure 12 we investigate if the treatment effect estimate is affected due to possible compositional changes. The center panel presents the estimates when the study sample is restricted to include only the individuals who are observed throughout the entire selected window. For the smallest window, the restricted sample is the same as the baseline sample and therefore the results are identical. For wider windows the restricted sample differs more from the baseline sample. For window sizes up to about 18 months, this has little impact on the treatment effect estimate. For wider windows, the effects are generally underestimated in the restricted samples. However, the general pattern of larger treatment effects in higher wage groups persists until window sizes of about 36 months.

As an additional check, the right panel of Figure 12 uses information from the full sample but constructs different samples by weighting the observation of an individual inversely proportionally to the number of observations of that individual within the window. Therefore, if there are observations missing in the window for an individual, this is compensated by giving more weight to the available observations. Overall, the results are very similar to the baseline results. This suggests that changes in the sample composition over event time have little impact on the estimation of the treatment effect.

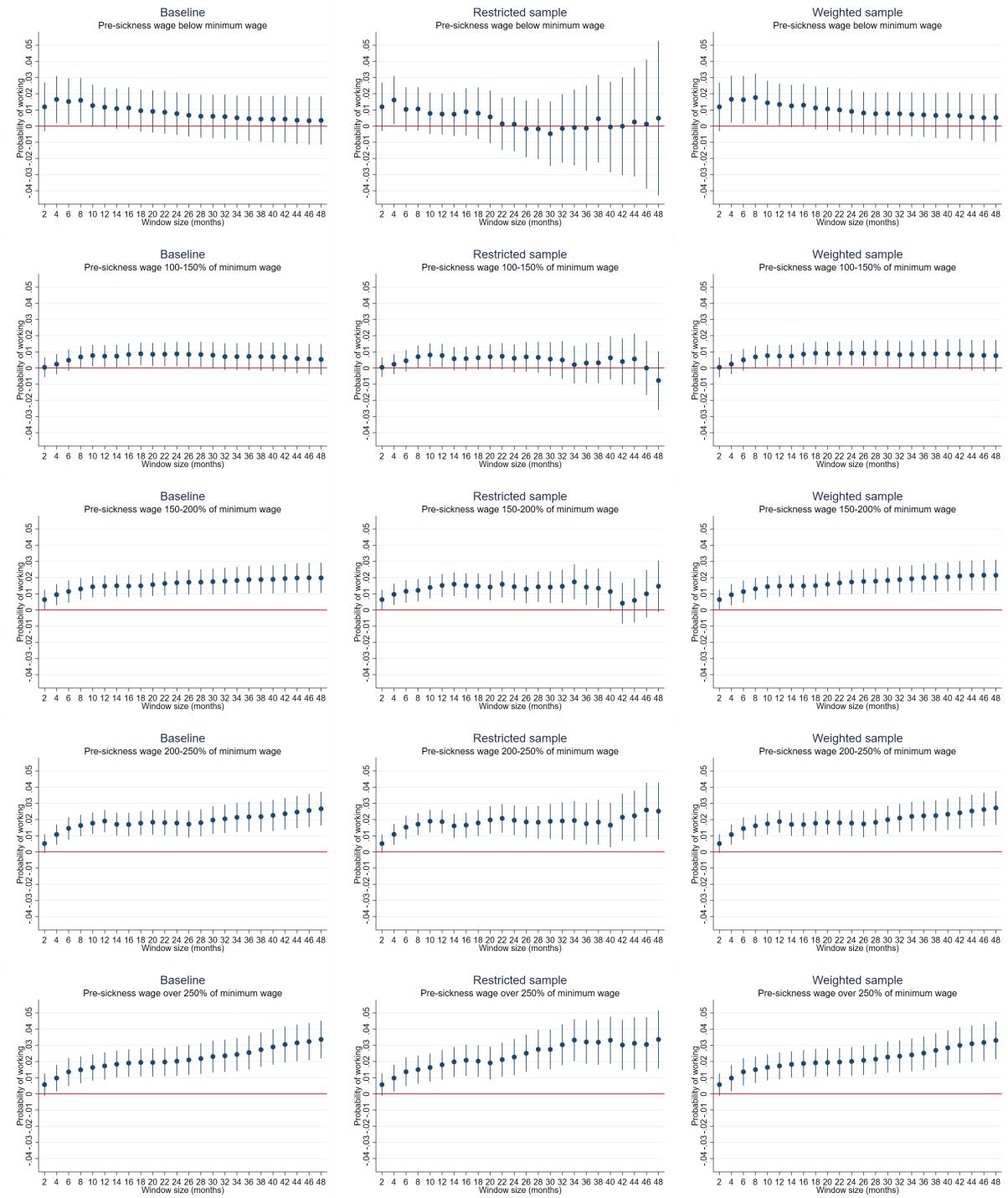


Figure 12: Treatment effect estimates using the baseline sample (left panel, reproducing Figure 7, restricted sample (center panel), and weighted sample (right panel). For each window, the restricted sample only includes individuals who are observed in the entire window. Observations in the weighted sample are weighted inversely proportionally to the number of observations of the individual in the window.

7.6 Continuity-based approach to RD

We consider the widely implemented continuity-based approach to regression discontinuity. Following standard practice, we consider a linear fit of the running variable and a triangular kernel for weighting the observations centered around the cutoff. We select the window which minimizes the mean squared error (MSE) of the RD estimator. In addition, for inference, we perform the same estimation in a window which minimizes the coverage error (CER) of the confidence interval. Table 4 presents the estimation results. The estimates mostly show an increasing effect of the financial incentives through higher wage groups, in line with the results based on the LR design in Table 2. However, none of the estimates are statistically significant. As explained in Section 5.1, use of the continuity-based RD design is questionable given the discrete nature of the running variable and the small number of distinct values available for the running variable. This casts doubt on the validity of the results based on this design.

Table 4: Results based on continuity-based RD

	Labor participation
Pre-sickness wage <100% of min. wage	0.008 (0.028)
Pre-sickness wage 100–150% of min. wage	0.009 (0.014)
Pre-sickness wage 150–200% of min. wage	0.008 (0.013)
Pre-sickness wage 200–250% of min. wage	0.011 (0.014)
Pre-sickness wage >250% of min. wage	0.013 (0.014)

Notes: For each wage group, separate optimal bandwidths are estimated at the left and right of the cutoff. The bandwidths are 19.7, 14.8, 20.6, 19.8 and 18.5 at the left of the cutoff, respectively from the lowest to the highest wage group. They are 19.1, 21.3, 14.4, 15.2 and 15.1 at the right of the cutoff. Bandwidths are chosen to minimize the mean squared error of the point estimate. Similar results are obtained when bandwidths are chosen to minimize the coverage error of the confidence interval. Regressions include the same covariates as in the baseline regression based on the LR approach. Robust bias-corrected standard errors are in parentheses.

7.7 Duration of DI receipt

In the baseline specification, to account for the possible effect of the duration of DI receipt, we included in the set of controls dummies each controlling a multiple of six months of benefit receipt duration. A stricter measure of the elapsed duration could affect the treatment effect if it is closely correlated with the running variable within a window size around the cutoff. The right panel of Figure 13 presents results when in the regression we allow for dummies controlling a multiple of three months of DI receipt duration. Compared to the baseline results in the left panel of the figure, there are two main differences. First, the magnitude of the treatment effect is most of the time smaller in all income groups. Second, the stricter measure of elapsed duration pronounces our main finding that the work resumption program has heterogeneous effects on income groups. While the two lowest income groups show no significant response to the program, higher income groups show significant responses for almost all different window sizes around the cutoff.

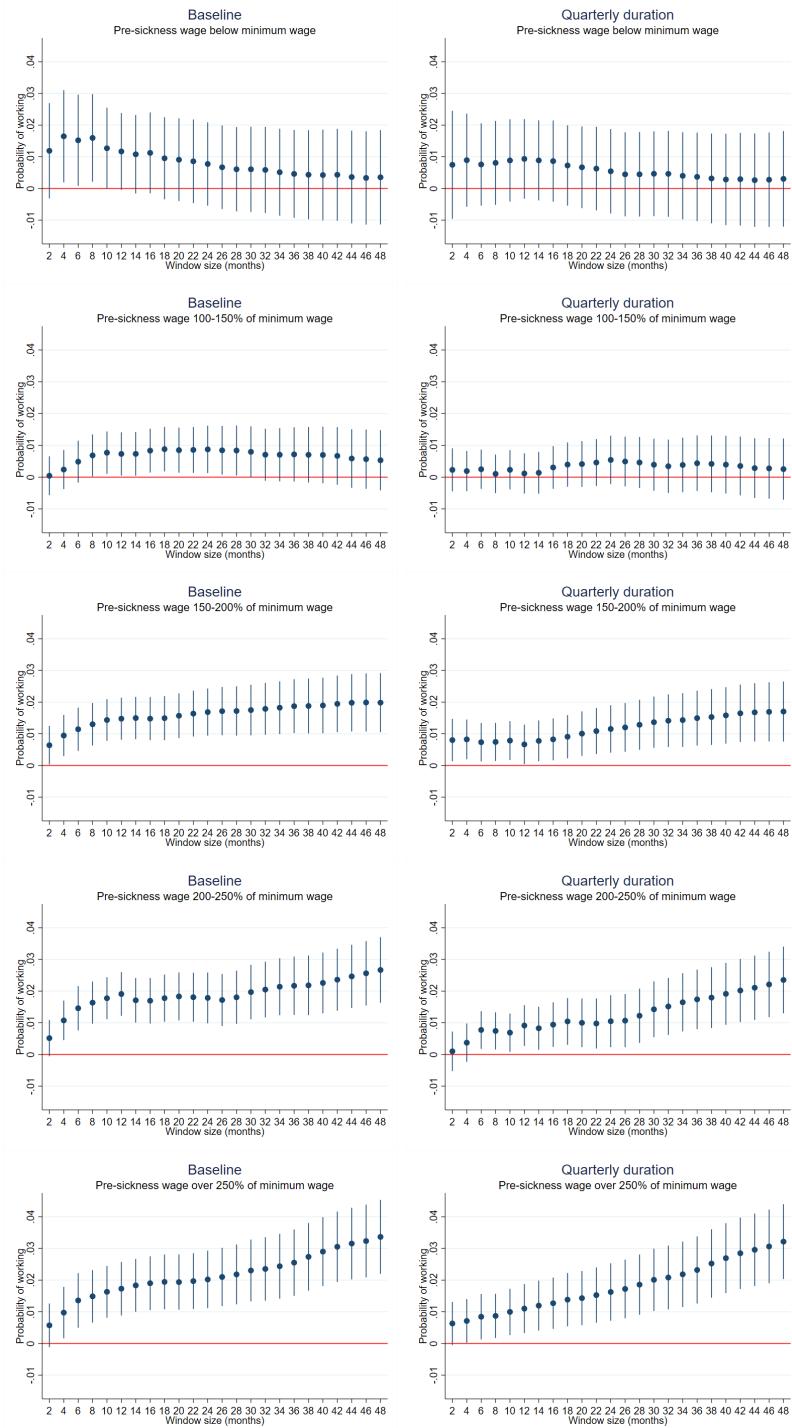


Figure 13: Treatment effect estimates based on regression specifications including dummies that control for multiples of six (left panel, reproducing results in Figure 7 and three (right panel) months of DI receipt duration.

7.8 Effects of the work resumption program by age

As explained in Section 5.3, we cannot control for age in our regression model. Instead, we analyze the heterogenous effects of the treatment with respect to age. Figure 14 presents results for young, middle and old aged partially disabled individuals. There are three main findings. First, there is a clear age effect. For younger individuals the impact of the program is stronger. Middle aged individuals respond only in the highest wage groups. This compares to [Koning and van Sonsbeek \(2017\)](#) who find no labor participation effect for this age group. Second, regardless of age, the impact of the program is stronger for individuals with higher pre-sickness wages. This is in line with the heterogenous incentives of the program that are stronger for individuals with higher pre-sickness wages. Third, within the age groups, the impact of the program is almost always very stable across larger window sizes providing supporting evidence for the identifying exclusion restriction assumption.

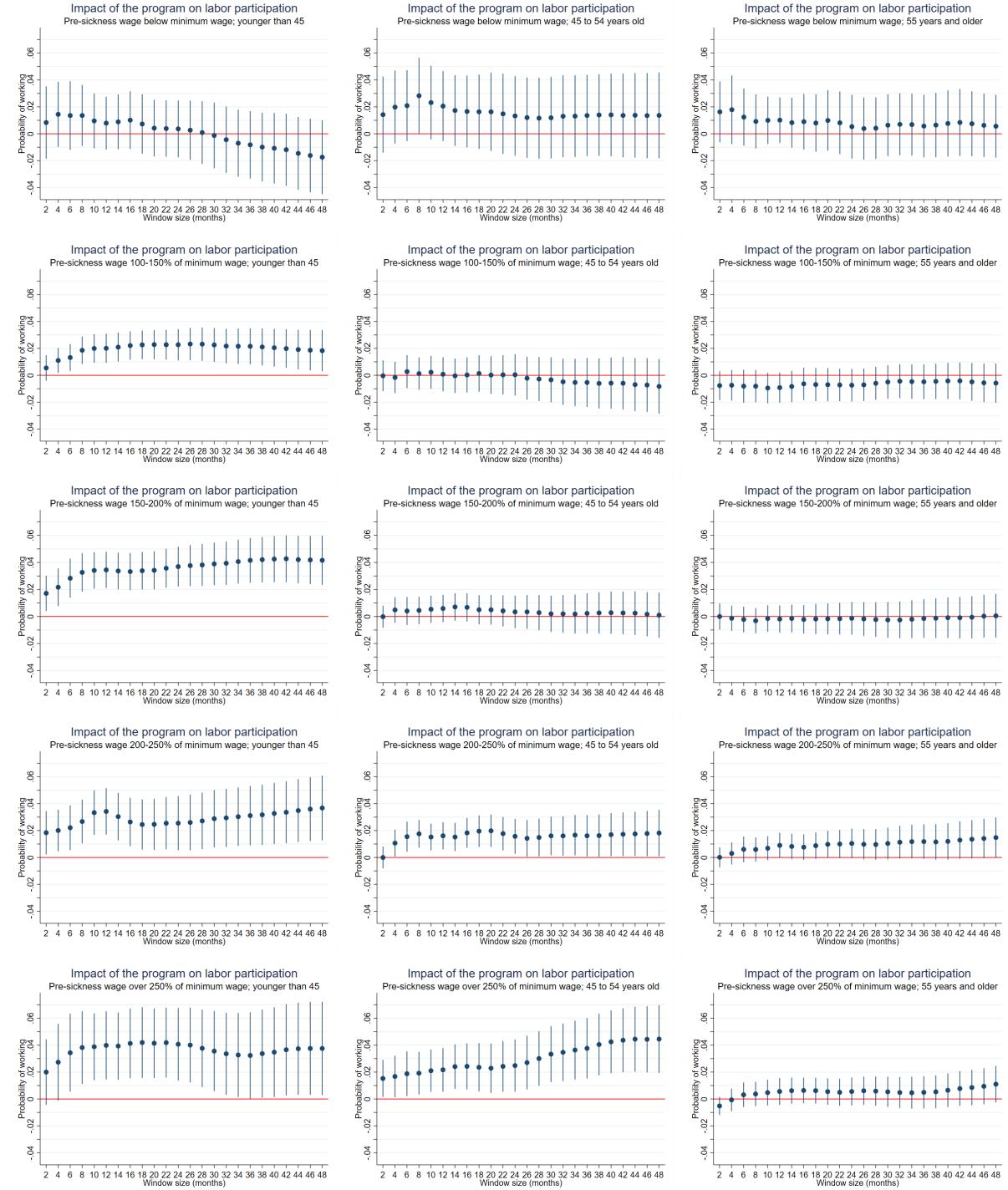


Figure 14: Treatment effect estimates for individuals who are younger than 45 years (left panel), 45 to 54 years old (center panel), and older than 54 years (right panel).

8 Conclusion

We analyzed the effects of the financial incentives of the work resumption program built into the DI scheme for partially disabled individuals. As the program incentives are larger for individuals who earn higher wages before they fall sick, we analyzed the effects of the program across earnings groups. We studied the program for its immediate and long-term effects. In line with the heterogeneous incentives of the program, we find substantial differences across pre-sickness wage groups.

Policy implications and recommendations of our study are four-fold. First, the main gains of the DI reforms in the past two decades have been reached by stricter screening and tighter eligibility criteria (Section 1). Our results suggest that financial incentives built into the DI scheme add to this significantly.

Second, the work resumption program is effective for individuals at the middle to higher end of the earnings distribution. This implies that the financial incentives of the work resumption program cannot be relaxed for these individuals without causing adverse effects in terms of lower labor participation and earnings. On the other hand, financial incentives are ineffective for more than a quarter of the participants of the work resumption program who are low-wage earners. Their incentive to resume working is comparatively low, and, on top of their health impairment, lack of skills and job opportunities may put additional constraints on their labor participation. Since benefits are already near the social minimum level for a large part of this group, the possibilities for increasing the financial incentives for work seem limited.

Third, the disparity in how income groups respond to the program suggests that the program might unintentionally increase inequality. Although the higher wage groups have more to lose from the strong financial incentives of the programme, they are also better able to compensate a disability related income loss through alternative income sources.

Finally, a policy recommendation regards how DI beneficiaries are informed of the incentives of the work resumption program. As explained in Section 2, a beneficiary receives two letters which inform about the program incentives; one at the moment the benefit is awarded and another shortly before expiration of the wage-related benefit. However, neither of the two letters explicitly informs about the size of the incentive and the difference between the two types of second-stage benefits, nor they explain how working more, below and above the 50% threshold, affects the benefit level and the total income from wage and benefit. By making this information explicit, beneficiaries can better anticipate the incentives to come so that effectiveness of the incentives can be improved.

Our falsification analysis showed that a number of covariates related to disability and background characteristics are not balanced around the cutoff. Although we argued that this is not likely to affect our qualitative results, covariate balance can be achieved using a data preprocessing method such as entropy balancing ([Hainmueller, 2012](#)). This is left to future research.

Acknowledgments

This research is supported by Netspar under grant number LMVP 2019.01. Its contents are the sole responsibility of the authors. We thank UWV, and in particular Lucien Rondagh, Willy van den Berk, Carla van Deursen, and Roel Ydema, for providing the disability data. We thank Hans Bloemen, Ziwei Rao, Arthur van Soest, and participants of the Netspar Pension Day 2021 and the Netspar International Pension Workshop 2022 for their helpful comments and suggestions.

References

- Butler, B., Volkerink, M., Lung, C. L., 2019. Nieuwe conjunctuur indicator voorspelt afvlakking groei in 2019. *Economisch Statistische Berichten* 104 (4773).
- Bütler, M., Deuchert, E., Lechner, M., Staubli, S., Thiemann, P., 2015. Financial work incentives for disability benefit recipients: lessons from a randomised field experiment. *IZA Journal of Labor Policy* 4 (1), 1–18.
- Campolieti, M., Riddell, C., 2012. Disability policy and the labor market: evidence from a natural experiment in Canada, 1998–2006. *Journal of Public Economics* 96 (3-4), 306–316.
- Cattaneo, M. D., Frandsen, B. R., Titiunik, R., 2015. Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. senate. *Journal of Causal Inference* 3 (1), 1–24.
- Cattaneo, M. D., Idrobo, N., Titiunik, R., 2020. A Practical Introduction to Regression Discontinuity Designs: Foundations. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press.
- Hainmueller, J., 2012. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 20 (1), 25–46.
- Kantarcı, T., van Sonsbeek, J.-M., Zhang, Y., 2019. The impact of the disability insurance reform on work resumption and benefit substitution in the Netherlands. *Netspar Discussion Paper* 01/2019-013.
- Koning, P., Lindeboom, M., 2015. The rise and fall of disability insurance enrollment in the Netherlands. *Journal of Economic Perspectives* 29 (2), 151–172.
- Koning, P., van Sonsbeek, J.-M., 2017. Making disability work? The effects of financial incentives on partially disabled workers. *Labour Economics* 47, 202–215.
- Kostøl, A. R., Mogstad, M., 2014. How financial incentives induce disability insurance recipients to return to work. *American Economic Review* 104 (2), 624–655.
- Malani, A., Reif, J., 2015. Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics* 124, 1–17.
- Mattei, A., Mealli, F., 2017. Regression discontinuity designs as local randomized experiments. *Observational Studies* 3 (2), 156–173.
- OECD, 2021. Public spending on incapacity (indicator).
- Ruh, P., Staubli, S., 2019. Financial incentives and earnings of disability insurance recipients: evidence from a notch design. *American Economic Journal: Economic Policy* 11 (2), 269–300.
- Sales, A. C., Hansen, B. B., 2020. Limitless regression discontinuity. *Journal of Educational and Behavioral Statistics* 45 (2), 143174.
- Vall Castelló, J., 2017. What happens to the employment of disabled individuals when all financial disincentives to work are abolished? *Health Economics* 26 (S2), 158–174.
- Van Sonsbeek, J.-M., Gradus, R. H. J. M., 2013. Estimating the effects of recent disability reforms in the Netherlands. *Oxford Economic Papers* 65 (4), 832–855.
- Weathers, R. R., Hemmeter, J., 2011. The impact of changing financial work incentives on the earnings of social security disability insurance (ssdi) beneficiaries. *Journal of Policy Analysis and Management* 30 (4), 708–728.
- Zaresani, A., 2020. Adjustment cost and incentives to work: Evidence from a disability insurance program. *Journal of Public Economics* 188 (104223).

Appendices

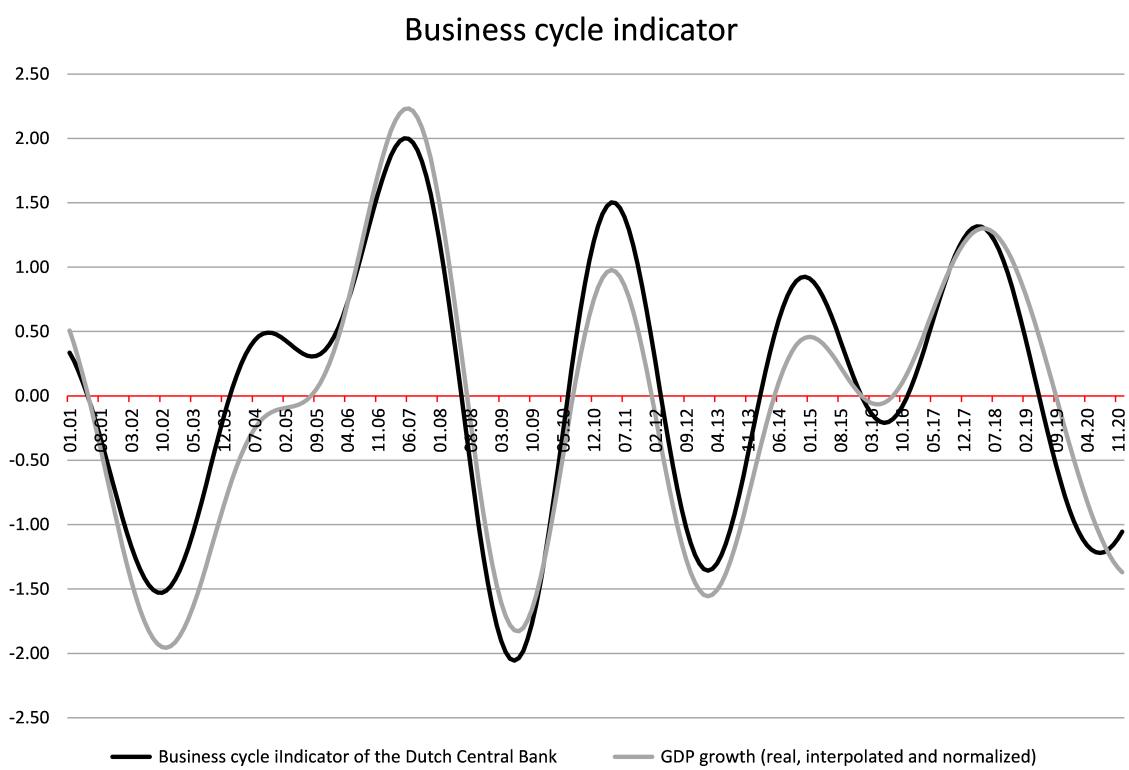


Figure 17: Business cycle indicator of the Dutch Central Bank.

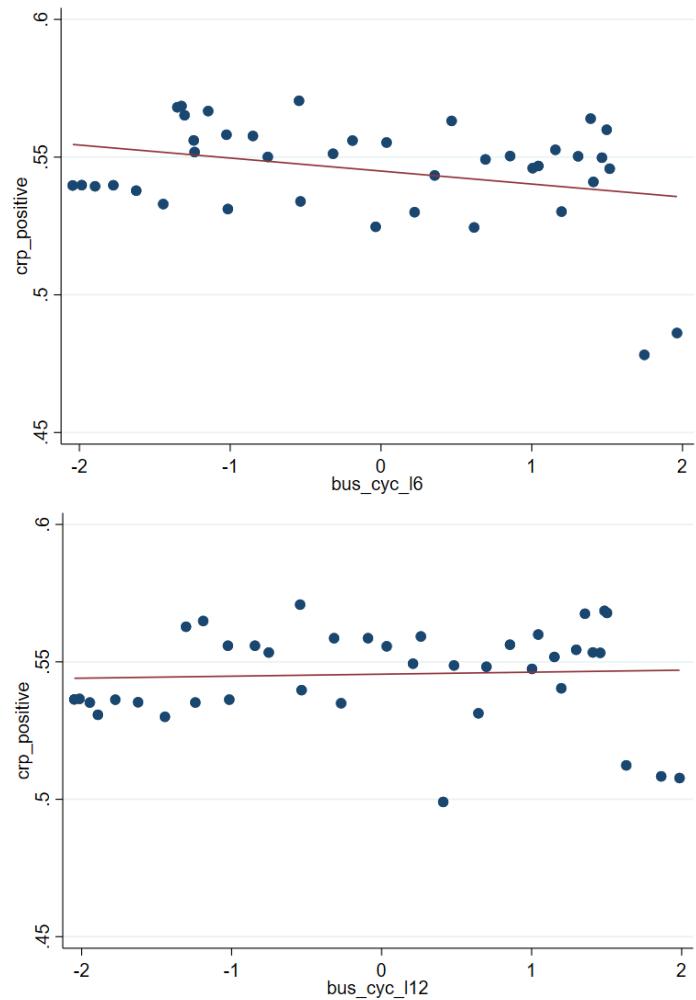


Figure 18: Binned scatter plot of mean labor participation against the business cycle indicator. Each dot represents the mean labour participation for that bin of the business cycle indicator. The fitted line is based on regression using the underlying data that make up the bins. The outcome is labor participation and the only covariate is the business cycle indicator. The top and bottom panels consider lags of 6 and 12 months of the business cycle indicator, respectively.